AFOSR 69-2059TR

AD700105

STUDIES IN THE SOCIAL PSYCHOLOGY OF STRESS

Department of Psychology

University of Illinois, Urbana, Illinois

MULTIPLE METHODS IN THE LONGITUDINAL STUDY OF SMALL GROUPS UNDER STRESS

John C. Stapert
Joseph E. McGrath

TECHNICAL REPORT
AUGUST 1969

Research on Social and Psychological Factors in Human Stress
Air Force Office of Scientific Research Contract AFOSR-AF-1161-67

Joseph E. McGrath Principal Investigator

1. This document has been approved for public release and sale; its distribution is unlimited.

Reproduced by the CLEARINGHOUSE for Federal Scientific & Technical Information Springfield Va. 22151

DDC RECEIVED FEB 6 1970 C

131

AFOSR 69-2059TR

STUDIES IN THE SOCIAL PSYCHOLOGY OF STRESS
Department of Psychology
University of Illinois
Urbana, Illinois

MULTIPLE METHODS IN THE LONGITUDINAL STUDY OF SMALL GROUPS UNDER STRESS

John C. Stapert
Joseph E. McGrath

August, 1969

Technical Report
Research on Social and Psychological Factors in Human Stress
Air Force Office of Scientific Research
Contract AFOSR-AF-1161-67
Joseph E. McGrath, Principal Investigator

1. This document has been approved for public release and sale; its distribution is unlimited.

ABSTRACT

This investigation was an attempt to assess the convergent and discriminant validity of 3 key variables (influence, affect, and activity) in small group research by conducting a systematic study which followed and extended the Campbell and Fiske multitrait-multimethod paradigm. 64 paid, volunteer, male undergraduates were divided into 18 3- and 4-man groups. During four weekly 2-hour sessions each group performed a series of comparable tasks, under varying task-load conditions (normal, overload, and underload). The 3 traits were measured repeatedly by each of 5 methods. Analyses disclosed widespread method variance among all rating methodologies. The implications of the findings for small group research and for research in social and psychological stress are discussed.

FORWARD

This research was conducted in conjunction with a program of research on Social and Psychological Factors in Stress (AFOSR Grant AF 1161-66 and AF 1161-67, J. E. McGrath, principal investigator).

It is an attempt to develop a methodology for systematic, repeated measurement of group process and performance variables over extended duration.

Other completed studies in the program include:

McGrath, J. E. (Ed.) Social and psychological factors in human stress. New York: Holt, Rinehart & Winston, (in press).

Steiner, I. D. Perceived freedom. In L. Berkowitz (Ed.) Advances in social psychology, Vol. V. New York: Academic Press, 1970.

Steiner, I. D. & Darroch, R. K. Relationship between the quality of counter-attitudinal performance and attitude change. Journal of Personality and Social Psychology, 1961, 4, 312-320.

Lowe, R. H. & McGrath, J. E. Predicting characteristics of group output from individual performance characteristics. T. R. Urbana, Ill.: AFOSR Contract AF 1161-67, Department of Psychology, University of Illinois, 1969 (in press).

TABLE OF CONTENTS

	Page
INTRODUCTION.	1
CHAPTER	
I: BACKGROUND OF THE RESEARCH.	5
II: EXPERIMENTAL PROCEDURES	27
III: RESULTS	38
IV: WILL THE REAL RELATIONSHIP PLEASE STAND UP?-- SOME EXERCISES IN INFERENCE FROM LIMITED EVIDENCE	89
V: FURTHER DISCUSSION AND SUMMARY.	99
LIST OF REFERENCES.	109
APPENDIX.	113

INTRODUCTION

Research on small groups, while extremely prolific in the last three decades, has produced a relatively low yield of cumulative empirical information. There are probably a number of reasons for this limited yield (cf. McGrath and Altman, 1966). Prominent among these reasons are two distinct but related methodological problems: (1) the lack of effort to establish construct validity of many key variables and to eliminate method variance in the measurement of these variables and (2) the lack of research on group process <u>through time</u>. These two problems are the foci of the present research.

The first of these two problems is a measurement problem. A major aspect of the validity of a measurement method depends upon how <u>well</u> its measurements correlate with other independent procedures used to measure the same trait (convergent validity). The validity of a measurement method also depends upon how <u>poorly</u> its measurements of one trait correlate with its measurements of other, different traits (discriminant validity). Thus, a measurement method can be invalidated either by (a) correlations with other methods of measuring the same trait that are too low for convergence or (b) correlation of the measurement of one trait with another trait that is too high for discrimination between traits.

Without both convergent and discriminant validity of measurement, it is not clear whether significant empirical findings are due to (a) actual significant relationships (trait variance) or to (b) methodology (method variance). What often happens is that different researchers

each use a different single method to measure a trait, and arrive at irreconcilable results. And in the face of the variety of results one cannot decide where trait variance and method variance have contributed to the significant findings.

This problem arises in many domains of psychological research. The experimental analysis of stress may be taken as a case in point. Of it, Appley and Trumbull (1967) report, "Unfortunately, one investigator relies on the GSR, a second on blood volume changes, a third on pulse rate or heart rate, a fourth on muscle action potential, and so on. In animals, the presence of stress is often inferred from feces counts, trembling, 'freezing' or washing behavior, and so on" (p. 6). They continue by citing the nearly incredible variety of findings which has resulted.

The second of the two problems, the lack of research through time, is a matter of the feasibility of longitudinal research. One problem has been the difficulty of securing subjects for repeated measurements over a period of time. Another has been the absence of any set of standardized tasks or activities which could be used over a series of occasions.

The approach here taken to the problem of convergence of measures and the elimination of method variance follows and extends the logic of the Campbell and Fiske (1959) multitrait-multimethod matrix. The approach to the study of group process through time represents a generalization of the Campbell and Fiske paradigm to include the third dimension of time-ordered occasions. In effect, the present study is an analysis of convergence and discrimination among a set of behavior variables (traits) measured by each of a set of methods, for the members of each of a set of groups. Each measurement was replicated over a set

of situationally distinct and time-ordered occasions.

Chapter I begins with a review of the methods of measurement which have been used in small group research and continues by reviewing the relevant longitudinal studies of small groups. Where possible these reviews concentrate on studies which have incorporated stress as an independent variable. Campbell and Fiske's multitrait-multimethod paradigm is then presented and extended to the present research.

A set of standardized tasks (Hackman, 1966) has recently been made available for social-psychological research of this type. These are described briefly and their usefulness for longitudinal studies of laboratory groups is noted. Chapter I concludes with a presentation of the variables selected for the present study, and the basis for their selection.

Experimental procedures occupy Chapter II. This includes descriptions of the experimental subjects, procedures followed in data-collection, and data-handling and scoring

Results are presented in Chapter III. For each measurement method used there are certain internal considerations (e.g., rater reliability, test homogeneity) and these are presented first. Convergent and discriminant validity of the measurement methods are then presented. Following this, the effects of time on the subjects are considered. Chapter III concludes with a presentation of the effects of stress on the behaviors of the subjects and on the products written by the experimental groups.

Chapter IV is a discussion of the experimental findings The discussion emphasizes findings which would easily have been overlooked or misinterpreted in smaller-scale, less systematic experimentation.

Chapter V is a discussion of the findings which emphasizes

the usefulness and potential of a systematic research strategy.

Supplementary forms, sample questionnaires, and some relevant technical materials, referred to in the text, may be found in the appendix.

CHAPTER I: BACKGROUND OF THE RESEARCH

A Survey of Methodologies Frequently Used in the Study of Small Groups

A variety of methodologies is available to the researcher who wishes to measure aspects of small group behavior. Among the more common are questionnaires and rating forms administered to group members, personality tests of various kinds, and observation systems. A consideration of these frequently-used methods reveals some weaknesses which deserve attention.

Ratings and Questionnaires

Rating forms and questionnaires have long been a favorite among social scientists. Webb, et. al. (1966) note that such self reports have been the dominant method of social science. And, while there are some standard questionnaires available, it is most common for each researcher to construct his own questionnaires and rating forms for use in his own research. This has resulted in an over-dependence upon non-standardized ratings and questionnaires.

Some of the weaknesses of ratings and questionnaires are obvious. First, it is likely that the subject has his own personal hypotheses about the nature and intent of the research in which he is participating (Dulany, 1968). And he will complete the ratings and questionnaires within the context of his private understanding of the experimental situation. Thus, the rating or questionnaire

items may have a sharply different meaning for him than was intended by the experimenter.

Second, a subject enters the experimental setting with his own response-strategy. The possibilities for bias here run all the way from stylistic "response set" biases to deliberate strategies on the part of the subjects. Subjects may wish to cooperate with the experimenter and help him confirm his hypotheses, they may wish to oppose the experimenter's interests and disconfirm the hypotheses, or they may have some other strategy (e.g., appearing "consistent" over a series of testing occasions). Whatever the strategy, intentional or unintentional, it may introduce a systematic but unspecified bias into the data.

In addition, it seems likely that most subjects will attempt to present themselves as favorably as possible. The concern over social desirability (see Scott, 1963) points to this difficulty. What "favorable" means to any particular subject may be difficult to specify, but if such a strategy operates, then ratings may contain systematic error from this source.

In his discussion of demand characteristics, Orne (1962) points out that an experimental subject often tolerates a considerable degree of boredom, discomfort, or actual pain during an experiment and therefore has a stake in the study in which he is participating. It follows that he must feel he has thereby made a useful contribution to scientific research. In part, this involves viewing himself as a "good subject." He will therefore behave in an experimental context in a manner designed to play the role of a "good subject," responding to questionnaires in a fashion designed to validate the experimental hypothesis as he sees it.

None of these weaknesses is fatal to the use of questionnaires and rating scales. In spite of the weaknesses, self reports remain

as a most useful measurement method. But the weaknesses are sources of possible systematic error which require checking against other methods.

Formal Tests

Formal tests are seldom used to measure group processes, as such, but they are often used as screening devices or pre-tests on the basis of which subjects may be grouped together. Among the usual advantages ascribed to tests are (1) they make it more difficult for a subject to be deceptive, (2) they are better standardized, and (3) their validity has been demonstrated by previous research. The extent to which these advantages hold for any particular test may be a matter of question, but they are features worth noting.

One drawback of using tests in the study of group processes is that tests are relatively inflexible with regard to the experimental situation. Behind most formal personality tests stands a set of assumptions about the traits being measured. Most often, it is assumed that the traits are relatively stable features which abide inside the person. It would be inconsistent with these assumptions to test repeatedly the same subjects on the same test, except for purposes of checking reliability. Changes in a subject's responses are typically viewed as a defect in the test, rather than as a change in the subject.

An advance in this area may be seen in the work of Leary (1957) and Schutz (1958, 1961) on interpersonal theories of personality. Interpersonal personality theories focus on interactions between people, rather than on inherent attributes of the person. Both Leary's and Schutz' theories have yielded personality tests which may provide sufficient flexibility for use in interpersonal situations.

Observational Systems

Several observational systems have been developed for use in

small group research (Bales, 1950; Lippitt, 1949; Murphy, 1937; Parten, 1933). The best-known and most widely used of these has been Bales' Interaction Process Analysis (IPA), which we shall consider briefly.

In the Bales system verbal interactions are said to fall into twelve categories: Shows Solidarity, Shows Tension Release, Agrees, Gives Suggestion, Gives Opinion, Gives Orientation, Asks for Orientation, Asks for Opinion, Asks for Suggestion, Disagrees, Shows Tension, and Shows Antagonism. An observer, often hidden from the group by one-way glass, monitors the verbal interaction and classifies, for each subject separately, the verbal interactions. The result is a time-ordered tabulation of each verbal unit, categorized by the twelve categories and by the speaker. By analyzing the frequencies for each subject, something about roles within the group can be learned; and by grouping the data into time intervals, group processes through time may be studied.

The great weakness of Interaction Process Analysis, and of other similar systems, is that it requires much interpretation of behavior by the observer as he is recording it. Verbal behaviors do not fall easily into their "proper" categories. And identical statements made in different contexts often fall into different categories. This means that the data generated by the system consist of <u>interpreted</u> behaviors of group members. And it may be worth noting that the interpretation must be done under unusually pressing circumstances, that is, during the group's interaction when rapid-fire exchanges are taking place among several group members and while each utterance is being colored by tone of voice, facial expression, and gestures.

At one level this gives rise to a problem of reliability between observers. For the difficulty of handling the category system under actual working conditions can foster unreliability. At a more basic level, however, it is a problem of validity, because all of the data must contain the systematic biases and errors introduced in the process of interpretation, even when multiple observers agree on the interpretation applied. And once the data have been recorded in the IPA format, there is no way of separating this methodological bias from the actual verbal and physical behaviors of the persons observed.

Other Methods

Webb, et. al. (1966) present a wide variety of classes of measures which might be used to supplement self-report methodologies and correct for systematic biases. While the methods suggested by Webb, et. al. may be more heuristic than elegant for most researchers, there are promising, unobtrusive, non-self-report methodologies which have been made available in recent years. Argyle and Kendon (1967) have reviewed a number of these techniques, some of which have been pursued in detail. The line-of-regard measure developed by R. Sommer (1959, 1961, 1962, 1965) in his research on interpersonal space and small group ecology is particularly noteworthy. Sommer has found that people adopt different seating patterns according to the type of interaction in which they are engaged. Seating arrangement, he suggests, is a major determinant of group process because of the way in which it facilitates or inhibits visual contact between people.

Summary and Implications

All methods of measurement contain error and each method tends to contain certain types of errors more than others. Furthermore, the

error tends to be systematic (i.e. correlated) over different traits measured by any given methodology. The tragedy is not, however, that the available methodologies contain systematic errors. It is, rather, that there has been a strong tendency to use a single methodology exclusively in each research project. As Webb, et. al. (1966) lament, there is an overdependence on a single, fallible methodology.

Any measurement taken on group activities may be partitioned into a component which reflects the trait purportedly measured and an error component. These have elsewhere been labeled "relevant" and "irrelevant" components (Campbell, 1960; Campbell and Fiske, 1959; Garner, 1954; Garner, Hake, and Eriksen, 1956). The error may be further partitioned into a number of subsections, including at least one segment of error which is peculiar to the method of measurement employed. Thus, if several methods were used to measure a particular trait, they would be expected to share a common relevant component and to differ in their patterns of irrelevant components. The implication is that each method of measurement should be checked against other methods having different patterns of irrelevant components (i.e., different sources of error).

For the most part, we have little information about the convergence of commonly-used measures of group processes. But in at least one case, some limited information is available. Hase and Goldberg (1967) compared four strategies of personality testing against two control strategies. Since the Hase and Goldberg study was confined to the general methodology of personality testing, thereby retaining the common systematic error due to this general methodology, one would have expected the validity coefficients to be relatively high. Unfortunately, it was found that even within the general methodology of formal personality testing, none of the

four strategies (factor analytic, contrasted groups, rational, and theoretical test-construction) correlated particularly well with the criterion when cross-validated, although all strategies did hold slight advantages over the control (stylistic and random test-construction) strategies. One implication of this study is that all strategies of testing introduce error variances, leaving the experimenter to choose the type of error with which he is most comfortable.

There are some more general implications however, which follow from this brief review of commonly-used methodologies. One implication is that multiple measures, subject to different sources of error, must be used in conjunction with each other more frequently. A second implication is that there is a general need to collect a body of systematic data, employing these commonly-used measurement methodologies, to enable investigation of convergence and differentiations of the measures with respect to each other. A third implication is that observational systems require developing in the direction of unobtrusive, objective (i.e., low inference) measurements. The strategy of the present study attempted to heed these implications.

The Study of Groups Through Time

It has long been recognized that the behaviors of group members change as time passes and the group establishes communication patterns, leadership roles, etc. The idea that group behaviors change as the group works through various phases in the process of solving problems goes back at least as far as John Dewey (1910), and has been studied more recently by Bales and Strodtbeck (1951). In order to identify the nature and pattern of these changes, groups must be studied through a period of time.

A number of researchers have called for studies in which a series of time-ordered observations is made (Appley and Trumbull, 1967; Firth, 1963; Leach, 1954; McGrath and Altman, 1966; Nadel, 1964; Smith, 1966). But in spite of the widely-recognized need, relatively few studies through time have been conducted, especially under well-controlled experimental conditions (Tuckman, 1965). This has resulted in inadequate understanding of group processes through time.

Time-related research on groups may be divided into two categories, depending upon the kinds of changing behavior patterns being investigated: (a) patterns of activity in groups which change from meeting to meeting, and (b) shifts or phase movements in activity within a single meeting. Research in the first category has consisted primarily of field studies, of which there have been only a few. Research in the second category has received some attention in the laboratory. Examples of each type will be cited briefly.

Shifts Between Sessions

Long-standing associations and long-duration missions have provided some opportunities for the study of a series of meetings of a single group. Newcomb (1961), in his well-known study of college men who lived together during the school year, dealt with changes in associations over time. His primary interest was in the formation and change of friendship relationships among his subjects. He found that friendship patterns among his subjects tended to stabilize during the first two weeks and to remain relatively unchanged thereafter.

Smith (1966), reports a study which he conducted while travelling with a seven-man Antarctic exploration group through its entire (four month) existence as it conducted geological studies on an ice field. Smith's chief psychological interests were in group structure and the

effects of personal danger, boredom, and monotony. Regarding group structure, he found that it took about a week for a work-structure to develop. In about three and one-half weeks an interpersonal structure developed. No structural changes were discovered in succeeding weeks.

Such field studies are especially valuable for identifying the important variables which may relate to group processes. Clearly, it seems on the basis of these studies that some aspects of group behavior do change over time. But further studies with stronger experimental control are necessary to isolate or refine such variables.

One developing body of research conducted by Haythorn, et. al. (Haythorn and Altman, 1967; Altman and Haythorn, 1965; Haythorn, Altman and Myers, 1966) has brought relatively long-duration (continuous 10-day observation) studies into the laboratory. The focus of Haythorn's research is the isolation of small groups and his primary concern is with the effects of isolation on behaviors of group members (e.g., self-disclosure). He has also investigated some effects of group composition (combinations of subjects having certain personality traits like dominance, dogmatism, etc.) on isolated groups. Unfortunately, this body of research is limited to studies of two-man groups. Although these studies are not constructed so as to assess the effects of time so much as the effects of isolation, they do recognize the need for psychological studies through time, and in part fill that need.

Somewhat related to the Haythorn studies is a body of research on stress in long-duration military groups. Much of this research has been conducted by Weybrew (1957, 1963). Related research has been conducted by Faucett and Newman (1953), Levine (1958), and Ritch (1948). An excellent review of these studies may be found in Weybrew (1967). This body of research is primarily concerned with the stress of long term isolated missions aboard submarines, and the dependent variables

measured include measures such as antialertness, critical flicker frequency, and perceptual spans. So it is difficult to assess, from these studies, the effects of the long-term missions on interaction, dominance in the group, and other interpersonal variables of interest in regard to group process.

R. F. Bales is one researcher who has attempted to study group interaction and its changes between successive sessions under controlled experimental conditions. Heinicke and Bales (1953) report an experiment which observed initially leaderless groups over a series of four meetings. Observation was by the Bales category system. It was found that members spent less time in task behavior and more in positive socio-emotional behavior as the series of meetings progressed. Negative socio-emotional behavior rose briefly during the second meeting, in what was labeled the "status struggle." A similar pattern was noted in a longer series of meetings in a study by Theodorson (1953). However, groups which met for only one session did not encounter the "status struggle" which typically took place during the second session of longer-duration groups.

In a study of ten 5-man groups which met for five consecutive weekly sessions, Deutsch (1949) found several significant differences between cooperative and competitive groups. Cooperative groups worked faster, were better coordinated, and had fewer communication difficulties. However, no striking developmental differences over time were found.

In summary, the research on group processes through time is relatively sparse and scattered. Some studies of long-duration missions tend to lack experimental control (as in the Antarctic studies). Others have not been primarily concerned with the effects of time or with the psychological dependent variables which would aid in understanding group process (studies of isolation and submarine studies). The work of Bales, et. al. has shown that significant changes in group

behavior may occur between sessions, although Deutsch failed to find between-session changes.

Phases Within Sessions

There has been some research on changes in group behaviors within a single meeting or session, although this subject area has barely been tapped. Bales has probably made as large a contribution here as any other researcher. In a large number of studies, Bales has identified three stages or phases (Orientation, Evaluation, and Control) through which problem-solving groups move within a single session (Bales, 1950a, 1950b; Bales and Gerbrands, 1948; Bales and Strodtbeck, 1964). Other research on within-session phases of activity have been relatively sparse. Tuckman (1965) has reviewed and integrated some 50 studies dealing with stages of group development over time, and has proposed a conceptual model for understanding observed stages. There are studies of therapy groups and their phases, but these seem to be unique to therapy groups and not easily generalizable to groups of other types (Talland, 1955).

Thus, some significant changes in group behavior which need further investigation seem to take place through time. Studies which consider only a single session of a group (whether that session be the first, last, or an intermediate one in a series), may be overlooking important changes in group behavior. Such studies also run the risk of projecting their empirical findings to groups "in general," which may be an unwarranted generalization. The implication which follows from this survey of research is that there is a strong need to collect a systematic body of data on group behaviors through time.

Campbell and Fiske's Multitrait-Multimethod Paradigm

Campbell and Fiske (1959) have detailed an approach to experimentation which allows [illegible] to separate trait variance from method variance and thereby to establish both convergent and discriminant validity for his methods. Briefly described, the Campbell and Fiske approach consists of measuring each of several traits by each of several methods. If the traits measured are functional unities and if the measurement methods are valid, the various methods should correlate well (converge) when measuring the same trait. The methods should correlate poorly (discriminate), both within a method and between methods, when measuring different traits.

The approach is illustrated in Table 1. The illustration involves three different traits, each measured by three different methods. The x-values in the matrix represent reliabilities. These could also be spoken of as monotrait-monomethod values. Campbell and Fiske intend test-retest reliabilities to be entered here, although there are other kinds of reliability which could be used. The y-values represent convergent validity, that is, the extent of agreement between two methods when used to measure the same trait. These may also be spoken of as monotrait-heteromethod values. High correlations here indicate that two methods attempting to measure the same trait are in agreement, thus lending credence to the belief that both are actually measuring the same trait.

The z-values are heterotrait-monomethod correlations, which represent method variance. When different traits are being measured, high heterotrait-monomethod values are a sign that a particular method agrees with itself regardless of the traits it is being used to measure. Low values, on the other hand, indicate that different traits

TABLE 1

A SYNTHETIC MULTITRAIT-MULTIMETHOD MATRIX

	Traits	Method 1			Method 2			Method 3		
		A_1	B_1	C_1	A_2	B_2	C_2	A_3	B_3	C_3
Method 1	A_1	x								
	B_1	z	x							
	C_1	z	z	x						
Method 2	A_2	y	w	w	x					
	B_2	w	y	w	z	x				
	C_2	w	w	y	z	z	x			
Method 3	A_3	y	w	w	y	w	w	x		
	B_3	w	y	w	w	y	w	z	x	
	C_3	w	w	y	w	w	y	z	z	x

are measured differently within the method. Thus, low values are desired here, which, in combination with high monotrait-heteromethod values, would indicate discriminant validity.

The w-values are heterotrait-heteromethod values. These are correlations between different traits as they are measured by different methods. Provided that reasonably valid measurement methods are being used, these values will indicate the true relationships among the traits.

The example in Table 1 could be extended to include as many traits or methods as one might desire. (The present research measured each of three traits by each of five methods.) It is important to consider the values in the multitrait-multimethod matrix in the order presented here. Since the validity of any measurement method is limited by its reliability, the reliabilities of the methods (x-values in Table 1) must be considered first. The other (y, z, and w) values must be considered in relationship to each other and, strictly speaking, need not be ordered. But it is convenient to look for the possibility of convergent validity early. If convergence among methods does not appear to be present (on the basis of the y-values), further consideration of the matrix is of little value. Given convergence, however, one may inspect the z-values for discrimination between traits.

As a final test of validity, the pattern of the w-values should be noted. Given reliability and validity of measurements, the w-values should represent the true relationships between pairs of traits. Since the underlying traits bear the same relationship to each other regardless of the measurement methods employed, the pattern of w-values within a heterotrait-heteromethod triangle should be the same, regardless of the heterotrait-heteromethod triangle being considered.

Second, there must be a set of methods which can be implemented simultaneously, or nearly so, within the situation in which the traits appear. The methods must be subject to different sources of error variance (see Webb, et. al., 1966). Ideally, of course, the methods would be completely independent of each other and therefore orthogonal with respect to error variance (see Garner, Hake and Eriksen, 1956).

Third, there must be a set of time-ordered occasions on which the measurement methods can be applied to the traits. These time-ordered occasions must be comparable to each other, so that the measurements may be meaningfully related with one another.

Hackman's Standardized Laboratory Tasks

The requirement of some basis for comparability among occasions in a time-ordered series has been a particularly difficult one for social-psychological research because of the lack of a set of standardized tasks or activities. Unless tasks are appropriately held constant, counterbalanced, or sampled throughout an experimental design, a real possibility exists that the results of the study may be seriously confounded with unintended task effects. Task type has proved to be a large factor in a number of studies. Hackman (1968) found that task type controlled well over 50% of the variance among group products in some dimensions.

Holding tasks constant over a series of occasions may place severe (though sometimes hazy) restrictions on the generalizability of results. And the repetition of the same task over occasions allows Ss to practice the tasks, thereby confounding practice effects with the effects of time. On the other hand, counterbalancing and sampling of task types encounter other problems: (1) Experimental manipulations are easily confounded with task variance and (2) the relevant parameters

of available tasks, necessary for appropriate counterbalancing and sampling, may not be known.

Fortunately, some recent research by Hackman (1966) has identified certain major parameters of written intellective tasks (task type and difficulty) and has provided a set of 108 tasks, balanced in terms of these parameters, which are appropriate for small-group experimental sessions. Hackman has also provided a set of scales along which products generated from these tasks may be evaluated.

The three task types isolated by Hackman are: production, discussion, and problem-solving. Production tasks require Ss to generate ideas about some topic. Discussion tasks require the selection of a "solution" from among a set of possibilities. Problem-solving tasks require Ss to decide how a particular solution can be executed. Each task requires about 15 minutes working time and each requires a written product. The tasks are graded into levels of difficulty within type. Normative data are available for small groups from size two to size seven, although the bulk of the data have been gathered from 3-man groups.

The availability of this set of standardized tasks has made it possible to use the multitrait-multimethod approach of Campbell and Fiske, testing each group over a series of occasions on different (but comparable) tasks. Thus, it has also become possible to manipulate situational factors during experimental sessions, since these manipulations may be kept orthogonal to the basic task parameters.

General Plan of the Study

The implications of the foregoing sections are for collection of a systematic body of data, having certain features, as follows: (1) The data must involve the measurement of a set of traits or behaviors

by each of a set of methods, (2) The traits must be sufficiently distinct from each other to allow the methods to discriminate between them, (3) The methods must be sufficiently independent to make them subject to different patterns of error variance, (4) The body of data must respect distinctions between types of group tasks, (5) Observations must be taken through time so that the effects of time may be noted. The following arrangement of experimental conditions was designed to produce a body of data conforming to these specifications.

Traits Measured

In a methodological study such as this, any of several sets of traits could have been selected. Those chosen for this particular study have been selected because they seemed fundamental to interpersonal process. Examination of a range of empirical and theoretical concepts suggests that three processes, influence, affect, and activity, are inherent in any interpersonal situation. The effects of many important member, group, and situational variables seem to be mediated through one or more of these three fundamental interpersonal processes (McGrath, 1963).

One source which identifies these three basic parameters is the research of Schutz (1958, 1961), who asserts that there are three basic interpersonal needs: inclusion, control, and affection. The inclusion dimension corresponds in part to activity. Control corresponds well to influence. And the affection dimension is self-evident.

Another source which identifies these three basic parameters, in slightly different, but parallel, terminology, is Osgood et. al. (1957). He has repeatedly found activity, potency, and evaluation to be dimensions of semantic space. While the contents of Osgood's and Schutz' work are quite different, there are marked similarities between matched pairs of concepts. In an interpersonal situation Schutz'

"inclusion" and Osgood's "activity" are both closely related to participation, or amount of interaction. Similarly, "control" and "potency" are both intimately tied to influence. And, in an interpersonal situation, Osgood's "evaluation" and Schutz' "affection" both have to do with the same parameter of valence or attraction between persons.

Hemphill (Hemphill and Westie, 1950) devised two dimensions of leader-member relations (initiation of structures, and consideration) which are similar to influence and affect. Other researchers who have found strikingly similar dimensions of interpersonal relations are Borgatta, Cottrell, and Mann (1958), Bales and Slater (1955), Carter (1951), Leary (1957), and Schaefer (1960).

These three dimensions also reflect the four primary dimensions of group structure as formulated by Cartwright and Zander (1960). Their concepts of the communication structure of a group, and its task structure or division of labor, both refer to patterns of interaction, with communication and task activity being specific forms of interaction. Their term "power structure" refers to the pattern of interpersonal influence. Their "sociometric" or "friendship structure" refers to the patterns of interpersonal affect or attraction.

One may reasonably conclude that the three parameters, influence, affect, and activity, are basic dimensions of any interpersonal situation. The present research attempted to measure these three parameters as the "behavior traits" of the study.

Methods of Measurement

The important feature of the methods of measurement used is that they differ in the sources of error variance to which they are subject. If error variance is confounded with trait variance, convergence

between methods could be due either to functional unity of the trait being measured or to error. But if the error variance in two (or more) methods differs, convergence between methods can more reasonably be attributed to functional unity of the trait being measured. Thus, a set of methods which differ in their sources of error must be chosen.

Five different methods of measurement were selected for this study: self reports, peer ratings, observer ratings, personality testing, and event records. The first three might be grouped together as a single, "rating" methodology. The implementation of these methodologies is detailed in Chapter II.

Occasions

The availability of Hackman's tasks allowed the present research to use the multitrait-multimethod approach of Campbell and Fiske, testing each group over a series of occasions on tasks which were different but comparable in type and difficulty. These standardized batteries of tasks also allowed the manipulation of a situational factor (stress) in a way that avoided confounding task variance with treatment variance.

The experimental groups met for four sessions of two hours each during which they worked on Hackman's tasks. The first hour of each of the four sessions followed the same procedure for all four sessions. This first hour of each session consisted of a normal, comfortable work load for the group. The second hour of each session varied from session to session. For the second hour of the first and third sessions the groups were subjected to "stress" in the form of a work overload. The second hour of the second and fourth sessions involved a "stressful" underload. Procedures for imposing the stress upon the groups are described in Chapter II. The second and fourth sessions were shortened

slightly to allow for administration of the Interpersonal Adjective Checklist (session two) and for de-briefing of the Ss at the end of the study (session four).

The experimental design is pictured in Figures 1 and 2. Figure 1 pictures the multitrait-multimethod matrix on the front "face" with occasions pictured in the third dimension. Figure 2 is a detailed picture of the experimental situations. Each of these figures exists conceptually in triplicate, one of each for each task type.

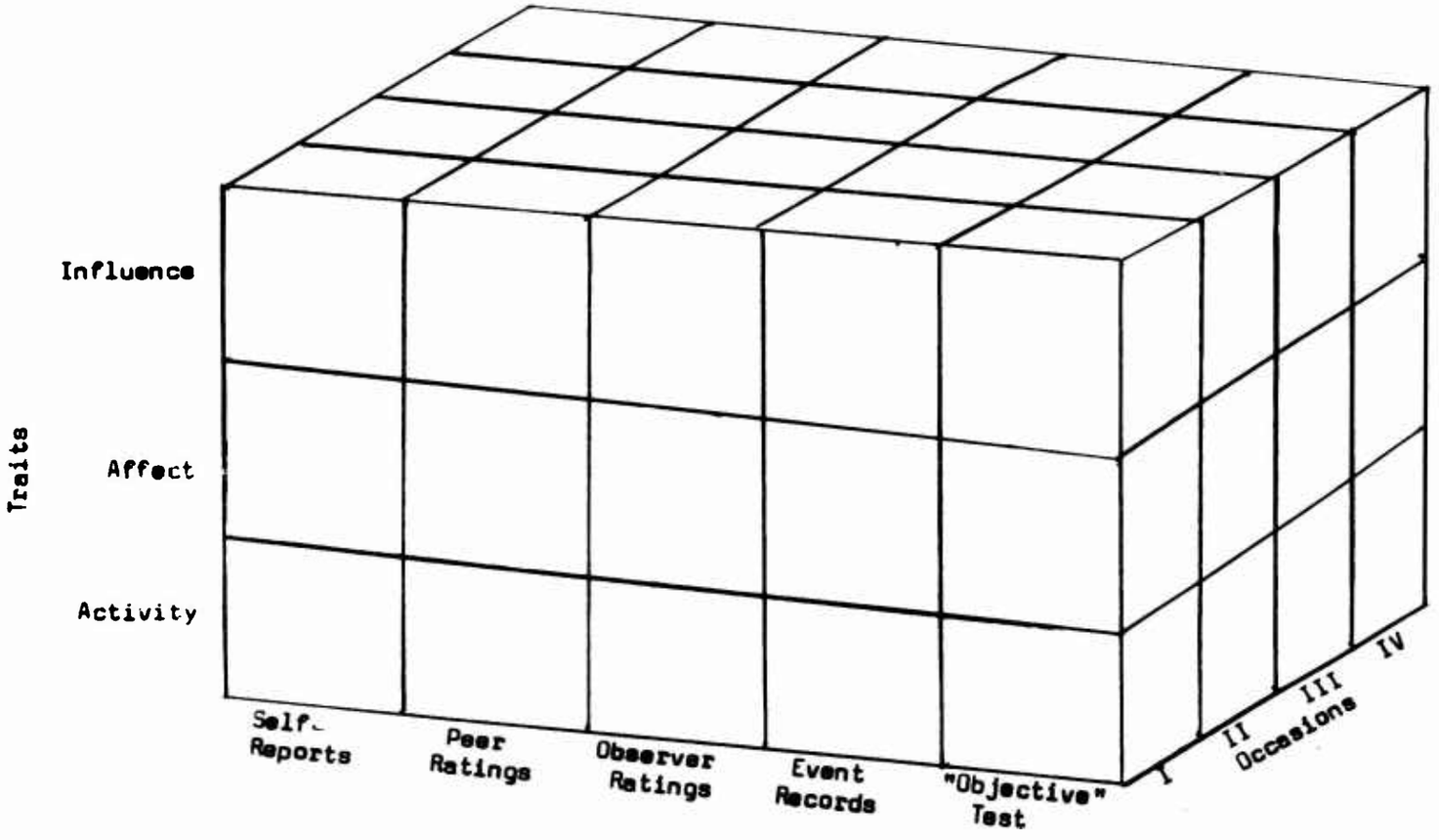


Figure 1. Over-all Design

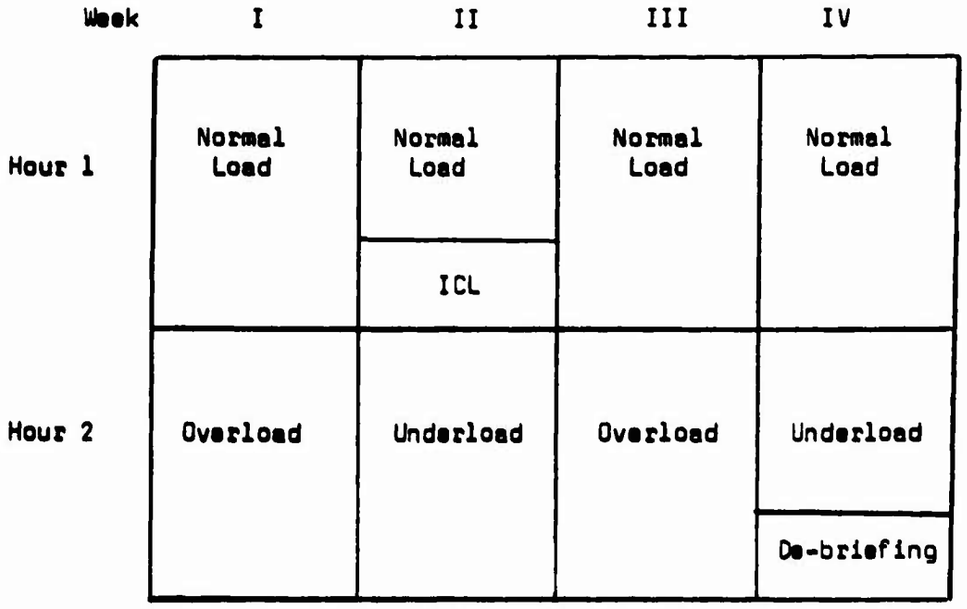


Figure 2. Experimental Sessions

CHAPTER II: EXPERIMENTAL PROCEDURES

Subjects and Data-Collection

The Ss were 64 male students in Psychology 100 at the University of Illinois. Ss were recruited for this study as they finished participation in Lowe's experiment.* Participation in Lowe's experiment was a course requirement, while participation in the present study was voluntary. Ss were offered $1.50 per hour for eight hours of work to participate in the present study.

Ss were divided among eighteen 3- and 4-man groups. Five groups (three 4-man and two 3-man) worked on production tasks. Six groups (two 4-man and four 3-man) worked on discussion tasks. And seven groups (five 4-man and two 3-man) worked on problem-solving tasks. The assignment of task type was based on the Ss' experiences in Lowe's study. Ss who had worked on production tasks in her study were assigned to production tasks in the present study, etc.

One of the 4-man discussion groups and one of the 4-man problem-solving groups lost a member after the initial session. One of the 4-man discussion groups and one of the 4-man problem-solving groups lost a member after session III. All other groups maintained their initial membership throughout the experiment. One of the 3-man discussion groups met for three sessions of two hours

---

*See appended note for the relationship of the present study to that of Rosemary Lowe.

each. All other groups met for four sessions of two hours each. The membership of the groups participating in the study in presented fully in Table 2.*

In general, the experimental sessions were held weekly and the four sessions were completed in four consecutive weeks. In some cases, however, adjustments in the schedule were made to accommodate Ss who had tests or were ill, and to allow for certain holidays and vacations.

During the experimental sessions Ss sat at a rectangular table located at one end of a long, narrow room. Two members sat along one side and one at each end (4-man groups). Ss in 3-man groups were free to choose any three of the four chairs, with most choosing the two end chairs and one along the back side.

An 8 mm movie camera was located at the other end of the room, and the experimenter sat at a table beside the camera. The camera was set on single frame mode and positioned so that the entire group appeared in the photographic field. The cameraman was signalled to photograph by a pair of Hunter decade timers, wired and adjusted to signal a five-second observation period at the beginning of each minute. Three photographs were taken during this observation period, spread evenly over the five seconds. While photographing, the cameraman noted which member(s) of the group was (were) talking. The instructions and record-keeping form for the cameraman may be found in the appendix.

At the beginning of each session each group was instructed to work at a normal, comfortable pace. The instructions were:

> During this first hour today we would like you to work at a normal, comfortable pace. The tasks are designed

---

*See appended note for a discussion of 3- vs. 4-man groups in this study.

TABLE 2

MEMBERS PRESENT PER GROUP PER SESSION

Group	Session			
	I	II	III	IV
11	4	4	4	4
13	4	4	4	4
14	3	3	3	3
16	4	4	4	4
17	3	3	3	3
21	4	4	4	3
22	3	3	3	-
23	3	3	3	3
24	3	3	3	3
25	3	3	3	3
26	4	3	3	3
31	3	3	3	3
32	4	3	3	3
33	4	4	4	3
34	4	4	4	4
35	4	4	4	4
36	3	3	3	3
37	4	4	4	4

> to take about 15 minutes each. So, you might expect to complete about three tasks during the first hour. But they do vary somewhat and different groups work at different speeds.
>
> Some groups work a little faster than others and some are more deliberate. So you might complete only one or two tasks, or you might finish four. In any case, work at a pace that is comfortable for this group. Signal me when you have finished a task and I will have some questionnaires for you to fill out.
>
> Here is your first task, you may begin.

At the completion of each task during this hour the experimenter administered the self-rating/peer-rating form to members of the group (appended form S-2-g). Data from this form were used to generate self-rating and peer-rating scores of influence, affect, and activity for each S. While Ss were completing the rating forms, the experimenter was rating all members of the group on the observer rating form (appended form 2-g). Data from this form were used to generate observer ratings of influence, affect, and activity for each S. Questions concerning motivation and coordination in the group were also included on the rating forms.

As soon as the rating forms were complete and collected, the group was instructed to begin the next task. A five-minute "break" was taken at the end of the first hour. The first hour of Session II was shortened by about five minutes in order to allow time for Ss to take the Interpersonal Adjective Checklist (LaForge, 1963; Leary, 1957).

The second hour of Sessions I and III was designed to produce an overload situation for the groups. They were told that information was needed about the maximum work load that groups could handle and therefore they should work as rapidly as possible. The instructions were:

> We need some information about the maximum work load which groups like this can handle. We would therefore like you

> to work as rapidly as possible without sacrificing the quality of your products. Keep the quality comparable to what you produced last hour.
>
> As an added incentive, we are offering a $5.00 per man bonus for the group which has the best over-all performance in this phase of the experiment. Your score from this hour will be combined with your scores from certain other hours in order to determine which group wins. For this particular hour it is speed that counts--the number of adequate products written. Here are the tasks for this hour. Do them in this order. When you finish one task, go right on to the next one. I will interrupt you at 15-minute intervals for rating questionnaires. You may begin.

The second hour of Sessions II and IV was designed to produce an underload condition. The instructions for these hours were:

> We need some information about the maximum quality which can be produced by groups of this type. So for this hour we would like you to write as complete and high quality a product as you can. Speed is of no importance here. You may well spend the full hour on one task, although there is a second task available if the first one does not completely fill the hour. I will interrupt you at 15-minute intervals for ratings. The $5.00 per man bonus applies to this hour. Here are the tasks. You may begin.

The subjects' and observers' rating forms used at 15-minute intervals during the second hour were identical to those used for the first hour.

The tasks used were those provided by Hackman (1966). In all, there were 108 different tasks, 36 in each of the three task types. A given group worked on tasks of only one type. The tasks in each type were randomly sequenced prior to the experiment. All groups within a task type experienced an identical sequence of these tasks.

Four tasks were provided for each of the normal load hours. Eight tasks were provided for each of the overload hours. And two tasks were available for each of the underload hours.

The data resulting from these procedures were: (1) up to

28 post-task questionnaires from each S and observer, (2) an Interpersonal Adjective Checklist from each S, (3) a series of photographs (of each group) which systematically time-sampled the task activities, and (4) up to 36 written products from each group.

Data Handling and Scoring

Post-Task Ratings

Each item on the questionnaires was scored on a seven-point scale, with seven representing the "high" end of the scale and one representing the "low" end of the scale. Influence was scored as the average of questions two and five. Affect was scored as the average of questions six and seven. And activity was scored as the average of questions two and three. Thus, each of the variables could range from one to seven for self, peer, and observer ratings. Questions one, four, eight, and nine were included so that there would be systematic relationships between Lowe's study and the present study. These questions were not scored for the present study.

The Interpersonal Adjective Checklists (ICL's)

The ICL's were scored in the standard way (Leary, 1957) in order to generate measures of influence (Leary's Dom-Sub scale) and Affect (Leary's Love-Hate scale). In addition to this standard scoring, the total number of items checked by S on the ICL was noted as an indicator of the S's activity level.

Event Records

The development of a method for noting and scoring behaviors of group members ("event records") may be one of the most useful results of the present research.

"Event records" were taken by means of photographing the group, on a time-sample basis, throughout the experimental sessions. These photographic records are supplemented only by the notation of the cameraman as to who was talking at the time he took the pictures. Scores of influence, affect, and activity were taken from the photographs as follows:

(a) Influence Given: A person influences others if he talks, if he is looked at by another person, and additionally if he is talking and is looking reciprocally at another person. Thus, a S received one "Influence Given" point for each time he talked, multiplied by the number of people to whom he talked (n-1). He also received an "Influence Given" point for being looked at by another S, and an additional point if, while he was talking, he and another S looked at each other reciprocally. These scores could vary from zero to nine.

(b) Influence Received: A person receives influence if he is talked to, if he looks at another person, and additionally if he looks reciprocally at another person who is talking. Thus, a person received an "Influence Received" point for each time he was talked to, one point for each person he looked at, and one additional point if he and a speaker were looking at each other. Influence received scores could vary from zero to nine.

(c) Positive Affect Given: A person exhibits positive affect when he smiles or laughs. Even if no one sees his smile, he is generating positive affect and "giving" it, as it were, to himself. If someone sees the smile (i.e., is looking at the smile), the positive affect is given to the person who sees it. Additionally, if the smiling person looks reciprocally at another person the positive affect is more strongly given. Thus, a subject received a "Positive Affect Given"

point for each time he smiled, another point for each S who looked at the smile, and one additional point if the subject was smiling and looking reciprocally at another S.

(d) Positive Affect Received: A person who smiles indicates that he is receiving the positive affect which he himself generates. A person who sees another person smile receives the positive affect from that other person. And there is additional positive affect received when a person looks reciprocally at another who is smiling. Thus a person received one "Positive Affect Received" point if he smiled, another if he watched someone else smile, and an additional point if he was looking reciprocally at another person who smiled.

(e) and (f) The logic and scoring of the Negative Affect Given and Negative Affect Received parallels that for positive affect. The only difference is the nature of the facial expression involved. Each affect score could range from zero to seven.

(g) and (h) Activity Level: Several behaviors were classed as activities, some of which overlap within the influence behaviors. The following behaviors received one point each as indicators of activity: talking (number of times), writing, moving of hands/arms, moving of body (posture or body orientation), and movement of head. For each person, two activity scores were computed (for each minute), a total activity score (as above) and a score of activity without talking. These scores could range from zero to seven.

This scoring of the data produced eight scores for each S for each minute of his group's activity. Thus, a serial, minute-by-minute record exists for each group on each of the three variables, influence, affect, and activity. The scores were appropriately adjusted to account for the numerical difference between 3- and 4-man

groups.[3] Sums of these minute-by-minute scores were combined for certain purposes of analysis, as described below.

For purposes of multimethod comparisons certain other scores were computed from the eight raw event records scores. Net influence was computed as influence given less influence received. These scores could range from nine to minus nine. Net affect was computed as positive affect given, less negative affect given (range: seven to minus seven). The activity score used in multimethod comparisons was total activity.

It will be noted that the scoring system detailed above differs substantially from the observation systems previously used in small group research. The above method is an objective system for noting observable behaviors and for scoring those behaviors as indicators of useful psychological concepts. Other systems of group observation require experimenters to interpret the behaviors as they observe them. The present system is considerably less inferential and offers the possibility of much more reliable measurement than other available group process analyses.

Products

Each hand-written product was typewritten and then rated on each of 20 scales by each of the three independent judges (after Hackman, 1966). The 20 scales represent six task-independent product dimensions (three rating scales each) and two task-dependent product dimensions (one rating scale each). A copy of the form on which the products were typed and scored may be found in the appendix.

---

[3] See appended note for a discussion of 3- vs. 4-man groups in this study.

The products were divided into two sets of about 300 products each. One set consisted of all products produced by the first nine experimental groups (three production type, three discussion, and three problem-solving), plus 19 products from Lowe's study. The other set consisted of all products from the other nine groups, and 82 different products from Lowe's study.

The rationale behind including some duplicates of Lowe's products in these product-sets was: Lowe's product sets consisted of many products, to each of only 12 tasks. This allowed for considerable variety of response within tasks, but relatively little variety between tasks. The sets of products generated by groups in the present study consisted of few responses to each of 108 tasks. This allowed for product sets with low variety of response within tasks, but high variety between tasks. It was thought worthwhile, especially for Lowe's research, to investigate whether or not these two different judging contexts affected the judgments made by the raters.

Raters were trained for their work according to the procedures developed by Hackman. Each judge was given a written description of the scale on which he was to rate a set of products. When the judge had read the description of the scale he sorted a set of sample products, for which criterion values were available. He then compared his judgments to the criterion judgments. Any discrepancies were discussed with one of the experimenters or with a more experienced judge until the discrepancies were satisfactorily resolved. Criterion values were not available for the sample products on the task-dependent scales so it was impossible to compare a judge's ratings with a criterion. However, all judges who rated on Adequacy or Creativity discussed the nature of the scale and several representative products with one of the

experimenters before proceeding with the ratings.

In all, 20 assistants were involved in the rating of products. Some judges rated products on several different scales, within the following limits: (1) No judge rated products on two scales in a single dimension, (2) No judge rated products on different scales in a single rating session, and (3) No judge rated products on any of the task-independent scales after having rated products on a task-dependent scale.

In general, it took a rater two and one-half to three hours to rate a set of products. In the cases of Adequacy and Creativity it often took five to six hours for a judge to rate a single set of products. This occasionally required that a judge take two sessions on consecutive days to rate a set of products on one of these task-dependent scales.

The three ratings given each product on each of the scales were averaged for each product and added to the averaged ratings for each of the two other highly correlated scales. This yielded eight scores for each product, six task-independent scores and two task-dependent scores. All scores could range from one to seven.

CHAPTER III: RESULTS

The amount of systematic data collected made a large number of analyses possible. The complete analysis of these data is a task which would consume several years. Many substantive, as well as methodological, questions could be asked of these data. A complete analysis, however, would obscure the major features of this research, burying the primary findings beneath numerous interesting, but secondary results.

In an effort to focus efficiently on the major questions involved in the research, certain main analyses were given priority and are presented in this chapter. These analyses speak to the measurement and methodological questions presented in the Introduction. Several supplementary analyses have been performed, but are reserved for brief mention in Chapter V.

This chapter begins with a brief presentation of some internal considerations of the measurement methods. These include appropriate aspects of measurement reliability. The correspondence of the Interpersonal Adjective Checklist to a circumplex model of personality is presented. For the products, the comparability of the present study to earlier work by Hackman is considered.

The main data analyses are rather global considerations of blocks of data. Convergent and discriminant validity of the measurement methods are examined for selected measurement occasions. The complete multitrait-multimethod matrices for eight occasions are presented.

The main effects of time of each of the three traits are presented as measured by each of the methods. Similarly, the main effects of stress on each of the three traits are presented as measured by each of the methods. The written products were also analyzed for effects of time and stress.

Internal Analyses of the Measurement Methods

Ratings

Pairs of questions were combined from the rating forms to yield measures of influence, affect, and activity. While no attempt was made to sample identical behaviors within each pair of questions, the relationship between questions in a pair is of interest.

The extent of agreement between any pair of questions varied from one measurement occasion to another. A systematic sample (viz.: for the first segment of the eight experimental hours) of the intercorrelations is presented in Table 3.

The two influence items correlated strongly for self ratings, peers, and observers. The strength of the correlation is lowest for self, next for peers, and highest for observers. The two affect items were very strongly correlated for all types of ratings, although somewhat less so for observers than for self or peer ratings. The two activity items were not as strongly related as the other pairs, although the mean correlations indicate significant positive relationships for all types of ratings. The activity items appear to have been more strongly related for the observers than for the subjects.

The differences in some of these correlations (between self, peers, and observers) seem to indicate that the rating scales had different meanings for different raters. This is most strikingly

TABLE 3

CORRELATIONS BETWEEN PAIRS OF
RATING QUESTIONS FOR SELECTED MEASUREMENT OCCASIONS

Occasion[a]	Trait[b]	Self			Peers			Observer		
		Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.
	N									
1	62	.58	.94	.33	.68	.93	.49	.66	.85	.64
5	62	.57	.93	.38	.62	.91	.57	.56	.64	.67
8	62	.65	.96	.32	.80	.95	.58	.78	.89	.77
12	62	.59	.94	.29	.67	.96	.52	.78	.81	.65
15	62	.54	.96	.39	.55	.96	.55	.65	.96	.57
19	62	.65	.98	.23	.79	.95	-.10	.64	.94	.38
22	57	.77	.99	.50	.76	.97	.58	.78	.86	.56
26	57	.50	.99	.49	.42	.94	.51	.79	.97	.58
Mean		.59	.96	.37	.66	.95	.46	.71	.87	.60

[a]These are the first segments of each hour.

[b]Key: Inf. = Influence, Aff. = Affect, Act. = Activity.

true in the case of the activity items. These differences exist in spite of the fact that the wording of the items and the instructions given were identical for all three types of ratings.

Interpersonal Adjective Checklist

(a) Conformity to the circumplex model: The Interpersonal Adjective Checklist is built on a circumplex model of personality (Leary, 1957). Figure 3 portrays the circumplex model with octants designated both by letter and by trait.

Table 4 presents the intercorrelations between octants of the Interpersonal Adjective Checklist generated by the present data. Adjacent octants did, in general, correlate as intended. The sole exception is FG-HI. Correlations between opposite octants are not strongly negative as the Interpersonal Adjective checklist model predicts. They tend, rather, to approach zero. Non-adjacent, non-opposite octants should be correlated at intermediate values, decreasing around the circle from adjacent to orthogonal to opposite octants and then increasing back to adjacent. This patterning is strong in Table 4. Only a few entries break this order, none of them by more than .05 units.

So, the octants fit a circumplex ordering, with one exception. However, opposite octants failed to generate negative correlations. One available explanation for this set of outcomes is that the "Submissive" (HI) octant and the octants adjacent to it are only infrequently endorsed by college males (LaForge, 1963). A similarly infrequent endorsement of submissive characteristics occurred in the present study. With low frequency of response in these octants there may be insufficient variability within these octants to allow strong correlations.

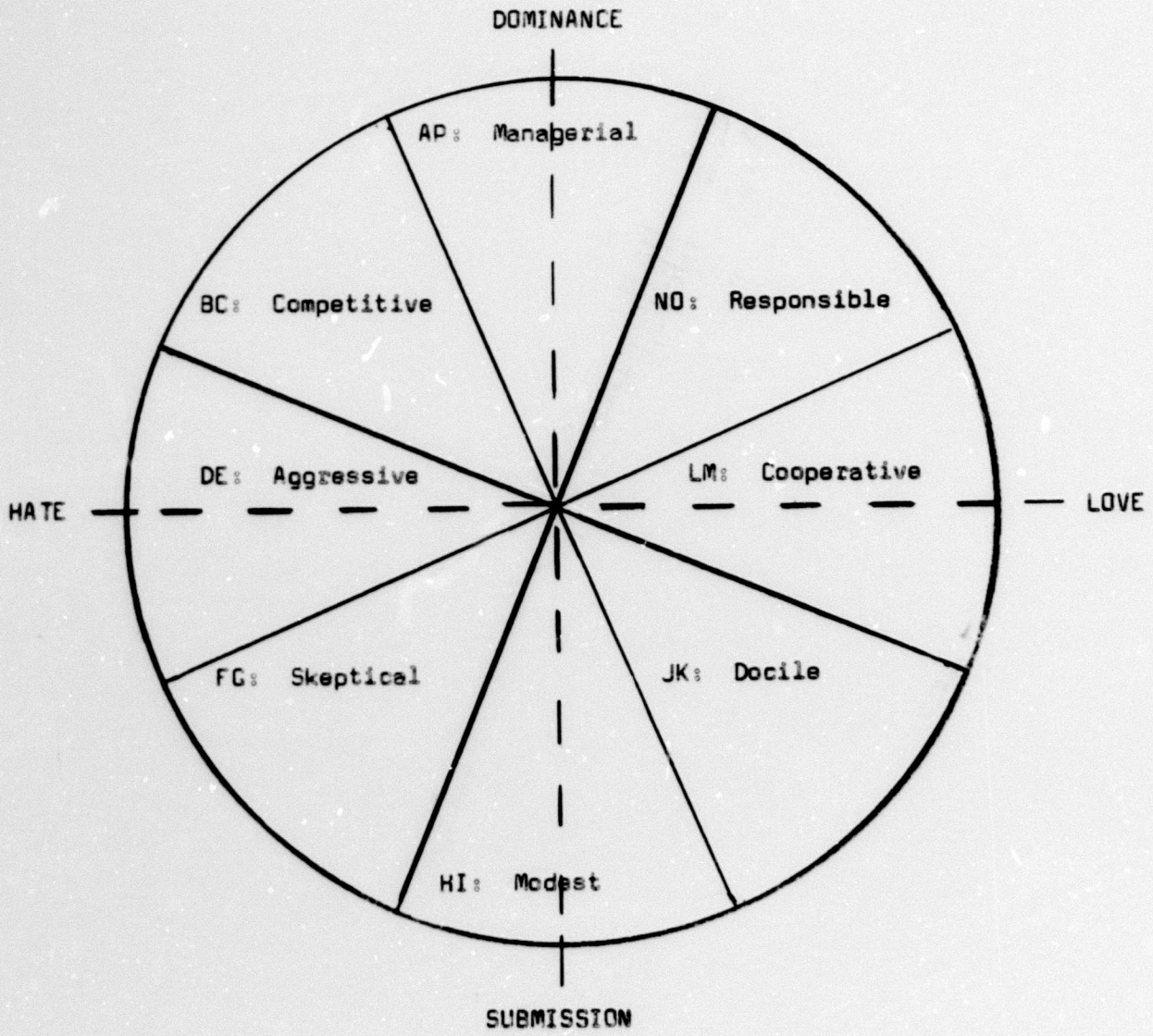


Figure 3. Circumplex Model of Personality (after Leary, 1957)

TABLE 4

INTERCORRELATIONS BETWEEN OCTANTS
OF THE INTERPERSONAL ADJECTIVE CHECKLIST
N=59

Octant	AP	BC	DE	FG	HI	JK	LM	NO
AP	1.00							
BC	.48	1.00						
DE	.35	.53	1.00					
FG	.20	.26	.51	1.00				
HI	.21	.01	-.05	-.02	1.00			
JK	.16	-.05	-.10	-.02	.51	1.00		
LM	.34	-.01	-.07	-.11	.33	.50	1.00	
NO	.39	.01	-.09	-.02	.30	.54	.68	1.00

(b) Reliability: Because of the interlocking relationship between the present study and that of Rosemary Lowe,* it was possible to examine the Interpersonal Adjective Checklist for both test-retest and split-half reliability. At the end of the Lowe study all Ss took a shortened, one-page form of the ICL which included half, although not a perfectly representative half, of the ICL items. This one-page form became the first page of the full two-page ICL administered to Ss who participated in the present study. Thus, test-retest reliability can be assessed from Ss' responses in the Lowe study vs. responses in the present study. Table 5 presents these test-retest reliabilities, by octant of the ICL.

Split-half reliability was calculated by comparing Ss responses on page one of the ICL with those on page two within the present study. Table 6 presents these split-half reliabilities.

*See appended note on the relationship of this study to that of Rosemary Lowe.

TABLE 5

TEST-RETEST RELIABILITIES OF THE ICL

AP	BC	DE	FG	HI	JK	LM	NO
.63	.47	.50	.58	.44	.61	.68	.60

TABLE 6

SPLIT-HALF RELIABILITIES OF THE ICL

AP	BC	DE	FG	HI	JK	LM	NO
.57	.52	.55	.46	.45	.24	.69	.63

The reliabilities were understandably attenuated by comparison with the published reliabilities of the ICL. Test-retest reliabilities were attenuated by (1) the two- to four-month lag between test and retest, an abnormally long interval for reliability testing, and (2) the fact that only half of the test could be used in this reliability calculation. Split-half reliabilities may have been attenuated because the two halves were not balanced. Thus, the reliabilities presented could well be considered as lower bounds of reliability for the octants of the ICL.

It must be remembered that any test, but especially a test based upon an interpersonal personality theory, is affected by the situation in which it is administered. This interaction of the test with the situation is customarily interpreted as unreliability. However, the test may in fact be reflecting the changing situation. This, and other possibilities, will be considered in Chapter V.

Event Records

For the event records, reliability may be considered virtually perfect. The events recorded were talking, looking at (line of regard), various body movements, and facial expression. Tabulations of talking were simply a matter of noting who was speaking and (later) counting the number of times each person talked during a 15-minute period. Given clerical accuracy, these values are numerically precise.

Looking at (line of regard), body movements, and facial expression were taken from the films. Experience and spot-checking demonstrated that independent film viewers could agree virtually perfectly on these variables. Once these events were recorded, the scoring system left no room for interpretation or judgment errors.

Products

Of major importance is the reliability among judges who rated the written products. The ratings of each rater on a set of products were intercorrelated. Correlations between pairs of raters were then pooled and transformed according to the Spearman-Brown prophecy formula to give a reliability estimate of the pooled judgments of the raters. The reliabilities of the pooled judgments are presented in Table 7.

The reliabilities range from .80 to .99, based upon six pairs of raters for each scale. These reliabilities indicate high agreement among raters and compare favorably with rater reliability in Hackman's earlier work in this area (Hackman, 1966).

Scores comparable to those in Hackman's study were achieved by Ss working under normal-load conditions. Table 8 presents Hackman's product scores for each task type and dimension. Parallel columns

TABLE 7

RELIABILITIES OF THE
POOLED JUDGMENTS OF PRODUCT RATERS
N=6

Dimension	Scale		
	I	II	III
Action Orientation	.97	.96	.92
Length	.99	- [a]	.97
Originality	.86	.91	.91
Outlook	.86	.83	.91
Quality of Presentation	.93	.87	.80
Involvement	.84	.87	.90
Adequacy [b]	.85		
Creativity [b]	.84		

[a] This scale is operational in [illegible].

[b] Adequacy and creativity are represented by only one scale each.

TABLE 8

COMPARISON OF MEAN PRODUCT SCORES
BETWEEN HACKMAN'S (1966) STUDY AND THE PRESENT RESEARCH

Dimensions	Means					
	Production		Discussion		Problem Solving	
	Hackman	Stapert	Hackman	Stapert	Hackman	Stapert
Action Orientation	2.75	2.34	3.95	3.43	5.89	6.13
Length	4.73	4.71	3.59	2.72	3.18	2.91
Originality	4.76	4.32	2.72	2.47	2.49	2.32
Outlook	3.88	3.97	4.13	3.93	4.64	4.32
Quality of Presentation	5.07	5.01	4.38	4.73	3.77	4.47
Involvement	3.06	2.23	5.39	4.88	3.93	3.62
Adequacy	5.11	4.94	4.97	4.39	5.10	4.90
Creativity	3.09	4.31	2.69	[illegible]	[illegible]	[illegible]

present mean product scores achieved under normal load conditions in the present study.

Inspecting Table 8 shows the consistency between Hackman's work and the present study. More striking than the proximity of paired mean values is the patterning of the means. This Table shows Action Orientation to be the characteristic scale for problem-solving tasks, Originality for production tasks, and Involvement for discussion tasks for both studies. All indications are, therefore, that ratings made on the products are reliable and that a high level of comparability exists between the present study and earlier work on these laboratory tasks.

Convergent and Discriminant Validity

Convergent and discriminant validity of the measurement methods may be examined in the multitrait-multimethod matrix. The matrix for the first measurement occasion (the first segment, first hour, first week, normal load)* is presented in Table 9. Several features of this matrix may be noted. (The major diagonal, reliability coefficients, of this and other matrices is omitted, since different types of reliability are relevant to different methods. Reliabilities are discussed in the earlier section of this chapter.)

The Interpersonal Adjective Checklist scores failed to show any pattern or correlate well with any other method. The failure of this test to converge with other methods was a consistent finding throughout

---

*"Segment refers to a period of time during which a group worked on a task and at the end of which ratings were taken. The length of a setment was variable during normal load hours, beginning and ending when a group began and ended a task. During the stress hours, segments were 15 minutes long.

TABLE 9

MULTITRAIT-MULTIMETHOD MATRIX
FOR OCCASION 1 N=62

	Self Ratings			Peer Ratings			Obs. Ratings			Test (ICL)			Event Records		
	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.
Self Ratings															
Influence															
Affect	(.45)														
Activity	(.83)	(.52)													
Peer Ratings															
Influence	.66	.17	.55												
Affect	.40	.15	.36	(.56)											
Activity	.60	.21	.48	(.86)	(.52)										
Obs. Ratings															
Influence	.45	.01	.29	.55	.41	.51									
Affect	.17	.22	.10	.03	.21	-.06	(.26)								
Activity	.38	-.01	.23	.44	.37	.43	(.85)	(.36)							
Test (ICL)															
Influence	.06	.07	.00	-.05	-.19	-.05	.06	.07	-.08						
Affect	-.30	.06	-.13	-.36	.00	-.31	-.10	.12	-.05	(-.03)					
Activity	.00	.01	.10	.03	-.04	-.13	-.09	-.05	-.21	(.34)	(.08)				
Event Records															
Influence	.31	-.23	.25	.27	.24	.24	.57	.05	.45	.19	-.02	-.01			
Affect	.30	.04	.31	.34	.31	.32	.12	-.05	.13	-.18	-.01	-.03	(.21)		
Activity	.28	.07	.29	.30	.31	.36	.59	.21	.52	.13	.03	-.05	(.60)	(.07)	

The following key applies to Tables 9 through 24:

KEY:

Underscored entries are convergent validity
Bracketed entries are (method variance)
Other entries are heterotrait-heteromethod values
Critical correlations for 60 df: $r \geq .25$, $p \leq .05$
$r \geq .33$, $p \leq .01$.

the several other multitrait-multimethod matrices, presented below. This is not surprising, since the relatively low reliabilities of the ICL, already noted, represent upper limits for all the correlations in the multitrait-multimethod matrix.

Convergent validity among the other methods ranges from near-zero to moderate. Convergence of methods for affect is very weak in most cases. Only peer ratings and event records show a significant multimethod convergence for affect, (.31).

Moderate convergence on influence and activity is shown between methods at several points. Self ratings, peer ratings, and observer ratings converge on these traits. But only the observer ratings show significant convergence with the event records. So, there seems to be convergence between some methods on some traits, but not between all pairs of methods nor on all traits.

If a measurement method is valid, it must not only converge with other measures of the same trait, but must discriminate between that trait and other traits. An inspection of the heterotrait-monomethod values in Table 9 shows that all three rating methods, especially self and peer ratings, fail to discriminate between traits.

Within each rating method, influence and activity are highly correlated. This is not surprising. The high correlations are partly artifactual, since "talking" counts both as influence and as activity.

In addition, correlations as high as .56 stand between influence and affect and between affect and activity within the rating methods. These high correlations are signals of method variance, generated when different traits are measured by a single method.

The Interpersonal Adjective Checklist and the event records

discriminated quite well between traits, except for a similar "built-in" correlation between activity and influence for event records. For the Checklist this has limited significance, since it failed to converge for any trait with any other method. But for the event records, which tend to converge with observer ratings for at least some traits, the discrimination is important.

Heterotrait-heteromethod correlations, which share neither trait nor method variance, are expected to be low. Some are low in Table 9, but others rise into the fifties. The high heterotrait-heteromethod values occur primarily between influence and activity and are concentrated in the rating methodologies. This may indicate a close relationship between influence and activity. It may also indicate the presence of method variance among rating-scales in general, in addition to method variance within specific ratings (e.g. self-ratings).

Multitrait-multimethod matrices for the first measurement occasion for each experimental hour are presented in Tables 10 through 16. Inspection of these tables indicates that, while specific values vary from table to table, the patterns noted in Table 9 seem to be quite consistent from one occasion to another.

The artifactually-inflated correlations between influence and activity, noted earlier, may have obscured convergence and discrimination among the methods. This artifact was removed for a subsequent analysis by using only question 2 on the rating forms as the measure of activity for those methods, and by using an activity score without talking for the event records. Tables 17 through 24 present the multitrait-multimethod matrices generated by this second analysis.

As expected, all correlations involving activity as a variable are lower than in the earlier analysis. Many of the correlations in

TABLE 10

MULTITRAIT-MULTIMETHOD MATRIX
FOR OCCASION 5 N=62

	Self Ratings			Peer Ratings			Obs. Ratings			Test (ICL)			Event Records		
	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.
Self Ratings															
Influence															
Affect	(.54)														
Activity	(.70)	(.48)													
Peer Ratings															
Influence	.50	.16	.37												
Affect	.34	.18	.41	(.52)											
Activity	.54	.23	.36	(.86)	(.49)										
Obs. Ratings															
Influence	.35	.02	.28	.53	.22	.47									
Affect	.27	.29	.19	.21	.45	.14	(.24)								
Activity	.39	.14	.35	.53	.35	.45	(.83)	(.23)							
Test (ICL)															
Influence	-.05	.13	-.06	.13	-.03	.05	-.04	-.06	-.08						
Affect	-.02	-.03	-.02	-.16	.08	-.11	-.13	.13	-.01	(-.03)					
Activity	.12	.11	.33	.07	.18	-.16	-.10	-.10	-.06	(.33)	(.07)				
Event Records															
Influence	.24	.00	.24	.44	.04	.30	.63	.03	.53	.13	-.14	.04			
Affect	.19	.14	.02	.16	.08	.24	.21	.08	.21	.03	.00	-.12	(.23)		
Activity	.24	.19	.23	.32	-.03	.19	.46	.13	.41	.15	-.14	-.02	(.54)	(.12)	

TABLE 11

MULTITRAIT-MULTIMETHOD MATRIX
FOR OCCASION 8 N=6.

	Self Ratings			Peer Ratings			Obs. Ratings			Test (ICL)			Event Records		
	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.
Self Ratings															
Influence															
Affect	(.43)														
Activity	(.79)	(.42)													
Peer Ratings															
Influence	.54	.20	.51												
Affect	-.03	.08	.13	(.39)											
Activity	.53	.18	.57	(.88)	(.41)										
Obs. Ratings															
Influence	.45	-.04	.38	.59	.05	.55									
Affect	.11	.11	.07	.19	.12	.21	(23)								
Activity	.42	.05	.42	.56	.19	.58	(.91)	(.21)							
Test (ICL)															
Influence	.00	-.03	-.11	-.09	-.10	-.13	.01	.08	-.07						
Affect	-.12	.13	.01	-.26	.06	-.14	-.32	.04	-.28	(-.03)					
Activity	.08	-.03	.16	-.09	-.11	-.15	.05	-.07	.02	(.33)	(.07)				
Event Records															
Influence	.28	.00	.25	.36	.08	.30	.53	.10	.50	.00	-.31	-.10			
Affect	-.02	.07	-.05	-.04	.07	.15	-.06	-.10	.01	-.08	.13	.01	(.09)		
Activity	.29	.07	.25	.23	-.06	.18	.24	-.04	.31	.14	-.16	.17	(.45)	(.07)	

TABLE 12

MULTITRAIT-MULTIMETHOD MATRIX
FOR OCCASION 12 N=62

	Self Ratings			Peer Ratings			Obs. Ratings			Test (ICL)			Event Records		
	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.
Self Ratings															
Influence															
Affect	(.33)														
Activity	(.64)	(.45)													
Peer Ratings															
Influence	.47	.25	.46												
Affect	-.04	.32	.29	(.35)											
Activity	.28	.30	.54	(.78)	(.49)										
Obs. Ratings															
Influence	.53	.12	.39	.56	.05	.51									
Affect	-.03	.14	.06	.01	.24	.20	(.14)								
Activity	.44	.14	.43	.57	.23	.58	(.88)	(.27)							
Test (ICL)															
Influence	.12	.07	.19	.17	-.02	.05	.01	.03	-.02						
Affect	-.34	-.14	-.14	-.31	.03	-.09	-.27	.17	-.20	(-.03)					
Activity	.04	.12	.23	.20	.11	.14	.04	-.07	.01	(.33)	(.07)				
Event Records															
Influence	.53	.07	.39	.52	.05	.39	.69	.01	.64	.15	-.41	-.04			
Affect	-.30	-.05	-.04	-.29	.01	.00	-.02	.10	-.03	-.09	.22	-.24	(.03)		
Activity	.31	.10	.37	.50	.18	.47	.50	.16	.56	.28	-.15	.27	(.65)	(-.02)	

TABLE 13

MULTITRAIT-MULTIMETHOD MATRIX
FOR OCCASION 15 N=62

	Self Ratings			Peer Ratings			Obs. Ratings			Test (ICL)			Event Records		
	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.
Self Ratings															
Influence															
Affect	(.39)														
Activity	(.72)	(.39)													
Peer Ratings															
Influence	.51	.25	.49												
Affect	.33	.21	.38	(.46)											
Activity	.37	.31	.47	(.84)	(.60)										
Obs. Ratings															
Influence	.34	.05	.18	.43	.07	.19									
Affect	.21	.15	.21	.00	.17	-.02	(.22)								
Activity	.42	.07	.33	.51	.07	.29	(.86)	(.16)							
Test (ICL)															
Influence	.05	-.03	.05	.15	-.01	.01	.17	.17	.07						
Affect	-.18	-.07	.02	-.05	.06	.01	-.10	-.08	-.06	(-.03)					
Activity	-.05	.05	.17	.12	.11	.02	.00	.23	-.04	(.33)	(.07)				
Event Records															
Influence	.34	-.04	.26	.40	.05	.20	.67	.04	.64	.20	.02	.09			
Affect	-.02	.18	.04	-.12	-.13	-.04	.01	.17	-.05	.15	-.03	-.09	(.11)		
Activity	.31	-.03	.26	.38	.11	.20	.63	.28	.62	.25	-.03	.31	(.70)	(-.05)	

TABLE 14

MULTITRAIT-MULTIMETHOD MATRIX
FOR OCCASION 19 N=62

	Self Ratings			Peer Ratings			Obs. Ratings			Test (ICL)			Event Records		
	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.
Self Ratings															
Influence															
Affect	(.52)														
Activity	(.66)	(.54)													
Peer Ratings															
Influence	.59	.20	.42												
Affect	.19	.11	.30	(.45)											
Activity	.43	.32	.46	(.65)	(.56)										
Obs. Ratings															
Influence	.50	.12	.36	.60	.20	.38									
Affect	.14	.17	.03	.20	.17	.07	(.22)								
Activity	.47	.06	.29	.53	.07	.35	(.84)	(.14)							
Test (ICL)															
Influence	-.05	.12	.08	.04	.02	-.21	.02	.05	-.04						
Affect	-.18	-.12	-.09	-.10	.09	.06	-.26	-.05	-.16	(-.03)					
Activity	.07	.09	.12	.17	.23	-.02	-.06	.11	-.03	(.33)	(.07)				
Event Records															
Influence	.17	.08	.22	.20	.00	.16	.63	-.03	.49	.03	-.23	-.06			
Affect	-.01	-.06	-.01	.05	-.01	.06	-.11	-.01	-.05	-.17	.12	-.31	(-.15)		
Activity	.33	.18	.21	.19	.03	.22	.45	.07	.52	-.01	-.04	-.04	(.58)	(-.17)	

TABLE 15

MULTITRAIT-MULTIMETHOD MATRIX
FOR OCCASION 22 N=57

	Self Ratings			Peer Ratings			Obs. Ratings			Test (ICL)			Event Records		
	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.
Self Ratings															
Influence															
Affect	(.48)														
Activity	(.80)	(.58)													
Peer Ratings															
Influence	.70	.43	.59												
Affect	.26	.52	.42	(.46)											
Activity	.54	.57	.54	(.83)	(.67)										
Obs. Ratings															
Influence	.40	.19	.27	.58	.14	.49									
Affect	.22	.25	.30	.32	.30	.33	(.32)								
Activity	.48	.16	.36	.52	.14	.43	(.89)	(.25)							
Test (ICL)															
Influence	.19	.03	.11	.12	-.09	-.02	.09	.09	.08						
Affect	-.16	-.09	-.08	-.17	.09	.03	-.06	-.03	-.05	(-.10)					
Activity	-.01	.07	.06	-.04	.15	-.10	.01	.08	-.08	(.33)	(.07)				
Event Records															
Influence	.46	.08	.34	.48	-.03	.30	.64	-.02	.59	.17	-.09	-.01			
Affect	.18	.37	.17	.19	.36	.39	.24	.03	.16	-.11	.04	-.09	(.12)		
Activity	.49	.34	.44	.49	.32	.55	.55	.28	.55	.17	.25	-.08	(.45)	(.33)	

TABLE 16

MULTITRAIT-MULTIMETHOD MATRIX
FOR OCCASION 26 N=57

	Self Ratings			Peer Ratings			Obs. Ratings			Test (ICL)			Event Records		
	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.
Self Ratings															
Influence															
Affect	(.48)														
Activity	(.68)	(.51)													
Peer Ratings															
Influence	.56	.34	.48												
Affect	.24	.28	.36	(.42)											
Activity	.46	.47	.43	(.78)	(.43)										
Obs. Ratings															
Influence	.43	.09	.28	.37	.03	.37									
Affect	.00	.04	.22	.17	.00	.30	(.34)								
Activity	.44	.04	.36	.38	.11	.35	(.90)	(.42)							
Test (ICL)															
Influence	-.08	.18	-.06	.02	-.10	.01	-.01	.01	-.02						
Affect	-.19	-.01	.01	-.06	.09	.05	-.16	.07	-.11	(-.10)					
Activity	.15	.12	.27	.13	.20	-.03	-.10	.03	-.08	(.33)	(.07)				
Event Records															
Influence	.39	.00	.24	.36	-.12	.30	.54	.07	.51	.04	-.13	-.07			
Affect	.01	.07	.10	-.02	.04	.13	.08	.32	.09	-.10	.05	-.02	(.18)		
Activity	.31	.13	.30	.40	-.01	.43	.43	.39	.47	.10	.08	.02	(.53)	(.33)	

TABLE 17

MULTITRAIT-MULTIMETHOD MATRIX
FOR OCCASION 1, SECOND ANALYSIS

	Self Ratings			Peer Ratings			Obs. Ratings			Test (ICL)			Event Records		
	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.
Self Ratings															
Influence															
Affect	(.45)														
Activity	(.47)	(.47)													
Peer Ratings															
Influence	.66	.17	.33												
Affect	.40	.15	.22	(.56)											
Activity	.33	.17	.19	(.51)	(.43)										
Obs. Ratings															
Influence	.45	.01	.02	.55	.41	.38									
Affect	.17	.22	.01	.03	.21	-.17	(.26)								
Activity	.28	-.01	.01	.35	.30	.23	(.62)	(.29)							
Test (ICL)															
Influence	.06	.07	.04	-.05	-.19	-.06	.06	.07	-.21						
Affect	-.30	.06	.04	-.36	.00	-.14	-.10	.12	-.02	(-.03)					
Activity	.00	.02	.12	.03	-.04	-.15	-.09	-.05	-.26	(.34)	(.08)				
Event Records															
Influence	.31	-.23	.06	.27	.24	.14	.57	.05	.27	.19	-.02	-.01			
Affect	.30	.04	.29	.34	.31	.27	.12	-.05	.12	-.07	-.01	.00	(.21)		
Activity	.15	.37	.13	.19	.16	.34	.34	.14	.32	.07	.02	-.02	(.28)	(.01)	

TABLE 18

MULTITRAIT-MULTIMETHOD MATRIX
FOR OCCASION 5, SECOND ANALYSIS

	Self Ratings			Peer Ratings			Obs. Ratings			Test (ICL)			Event Records		
	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.
Self Ratings															
Influence															
Affect	(.54)														
Activity	(.34)	(.35)													
Peer Ratings															
Influence	.50	.16	.19												
Affect	.34	.18	.35	(.52)											
Activity	.50	.26	.13	(.62)	(.41)										
Obs. Ratings															
Influence	.35	.02	.15	.53	.22	.35									
Affect	.27	.29	.16	.21	.45	.06	(.24)								
Activity	.36	.17	.24	.42	.31	.30	(.62)	(.17)							
Test (ICL)															
Influence	-.05	.13	.03	.13	-.03	.03	.04	-.06	-.07						
Affect	-.02	-.03	.02	-.16	.08	-.05	-.13	.13	.05	(-.03)					
Activity	.12	.11	.31	.07	.18	-.16	-.10	-.10	-.11	(.33)	(.07)				
Event Records															
Influence	.24	.00	.17	.44	.04	.16	.63	.03	.41	.13	-.14	.04			
Affect	.19	.14	-.06	.16	.08	.32	.21	.08	.21	.03	.00	-.12	(.23)		
Activity	.11	.16	.15	.14	-.14	.04	.20	.03	.07	.17	-.13	-.01	(.29)	(.03)	

TABLE 19

MULTITRAIT-MULTIMETHOD MATRIX
FOR OCCASION 8, SECOND ANALYSIS

	Self Ratings			Peer Ratings			Obs. Ratings			Test (ICL)			Event Records		
	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.
Self Ratings															
Influence															
Affect	(.43)														
Activity	(.36)	(.29)													
Peer Ratings															
Influence	.54	.20	.31												
Affect	-.03	.08	.26	(.39)											
Activity	.43	.10	.44	(.56)	(.36)										
Obs. Ratings															
Influence	.45	-.04	.13	.59	.05	.36									
Affect	.11	.11	.04	.19	.12	.13	(.23)								
Activity	.33	.08	.33	.49	.28	.48	(.74)	(.17)							
Test (ICL)															
Influence	.00	-.03	-.03	-.08	-.10	-.12	.01	.08	-.09						
Affect	-.12	.13	.12	-.26	.06	-.03	-.32	.04	-.14	(-.03)					
Activity	.08	-.03	.16	-.09	-.11	-.09	.05	-.07	.04	(.33)	(.07)				
Event Records															
Influence	.28	.00	.14	.36	.08	.12	.53	.10	.36	.00	-.31	-.10			
Affect	-.02	.07	-.03	-.04	.07	-.16	-.06	-.10	.09	-.08	.13	.01	(.09)		
Activity	.16	.11	.13	.06	-.06	.06	.07	.06	.14	.18	-.01	.23	(.14)	(.07)	

TABLE 20

MULTITRAIT-MULTIMETHOD MATRIX
FOR OCCASION 12, SECOND ANALYSIS

	Self Ratings			Peer Ratings			Obs. Ratings			Test (ICL)			Event Records		
	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.
Self Ratings															
Influence															
Affect	(.33)														
Activity	(.16)	(.35)													
Peer Ratings															
Influence	.47	.25	.20												
Affect	-.04	.32	.41	(.35)											
Activity	.06	.31	.47	(.45)	(.50)										
Obs. Ratings															
Influence	.53	.12	.10	.56	.05	30									
Affect	-.03	.14	.06	.01	.24	.27	(.14)								
Activity	.25	.12	.12	.46	.27	.44	(.60)	(.36)							
Test (ICL)															
Influence	.12	.07	.20	.17	-.02	-.02	.01	.03	-.04						
Affect	-.34	-.14	.08	-.31	.03	.04	-.27	.17	-.14	(-.03)					
Activity	.04	.12	.35	.20	.11	.07	.04	-.07	.02	(.33)	(.07)				
Event Records															
Influence	.53	.07	.13	.52	.05	.16	.69	.01	.49	.15	-.41	-.04			
Affect	-.30	-.05	.13	-.29	.01	.11	-.02	.10	-.02	-.09	.22	-.24	(.03)		
Activity	.12	.13	.25	.38	.23	.37	.22	.26	.38	.32	.02	.36	(.35)	(-.08)	

TABLE 21

MULTITRAIT-MULTIMETHOD MATRIX
FOR OCCASION 15, SECOND ANALYSIS

	Self Ratings			Peer Ratings			Obs. Ratings			Test (ICL)			Event Records		
	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.
Self Ratings															
Influence															
Affect	(.39)														
Activity	(.36)	(.32)													
Peer Ratings															
Influence	.51	.25	.34												
Affect	.33	.21	.39	(.46)											
Activity	.27	.30	.46	(.58)	(.65)										
Obs. Ratings															
Influence	.34	.05	.02	.43	.07	.02									
Affect	.21	.15	.18	.00	.17	.03	(.22)								
Activity	.34	.00	.23	.41	-.03	.14	(.57)	(.11)							
Test (ICL)															
Influence	.05	-.03	.02	.15	-.01	-.05	.17	.17	-.02						
Affect	-.18	-.07	.10	-.05	.06	.06	-.10	-.08	-.03	(-.03)					
Activity	-.05	.05	.29	.12	.11	.04	.00	.23	-.06	(.33)	(.07)				
Event Records															
Influence	.34	-.04	.11	.40	.05	.03	.67	.04	.43	.20	.02	.09			
Affect	-.02	.18	.04	-.12	-.13	-.03	.01	.17	-.07	.15	-.03	-.09	(.11)		
Activity	.19	-.10	.14	.31	.08	.05	.40	.17	.29	.24	-.07	.43	(.46)	(-.24)	

TABLE 22

MULTITRAIT-MULTIMETHOD MATRIX
FOR OCCASION 19, SECOND ANALYSIS

	Self Ratings			Peer Ratings			Obs. Ratings			Test (ICL)			Event Records		
	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.
Self Ratings															
Influence															
Affect	(.52)														
Activity	(.19)	(.38)													
Peer Ratings															
Influence	.59	.20	.10												
Affect	.19	.11	.29	(.45)											
Activity	.02	.25	.48	(-.04)	(.38)										
Obs. Ratings															
Influence	.50	.12	.08	.60	.20	-.02									
Affect	.14	.17	-.12	.20	.17	-.15	(.22)								
Activity	.29	.01	.00	.24	-.06	.10	(.37)	(.03)							
Test (ICL)															
Influence	-.05	.12	.08	.04	.02	-.26	.02	.05	-.13						
Affect	-.18	-.12	.01	-.10	.09	.15	-.26	-.05	-.08	(-.03)					
Activity	.07	.09	.09	.17	.23	-.10	-.06	.11	-.09	(.33)	(.07)				
Event Records															
Influence	.17	.08	.15	.20	.00	.08	.63	-.03	.12	.03	-.23	-.06			
Affect	-.01	-.06	-.02	.05	-.01	-.02	-.11	-.01	.07	-.17	.12	-.31	(-.15)		
Activity	.20	.18	.04	-.02	.02	.22	.14	.06	.35	-.01	.05	.03	(.30)	(-.22)	

TABLE 23

MULTITRAIT-MULTIMETHOD MATRIX
FOR OCCASION 23, SECOND ANALYSIS

	Self Ratings			Peer Ratings			Obs. Ratings			Test (ICL)			Event Records		
	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.
Self Ratings															
Influence															
Affect	(.48)														
Activity	(.47)	(.53)													
Peer Ratings															
Influence	.70	.43	.40												
Affect	.26	.52	.48	(.46)											
Activity	.31	.56	.39	(.57)	(.73)										
Obs. Ratings															
Influence	.40	.19	.14	.58	.14	.30									
Affect	.22	.25	.29	.32	.30	.27	(.32)								
Activity	.38	.01	.21	.27	.07	.13	(.59)	(.15)							
Test (ICL)															
Influence	.19	.03	.03	.12	-.09	-.09	.09	.09	.09						
Affect	-.16	-.09	.04	-.17	.09	.10	-.06	-.03	.10	(-.10)					
Activity	-.10	.07	.16	-.04	.15	-.05	.01	.08	-.09	(.33)	(.07)				
Event Records															
Influence	.46	.08	.18	.48	-.03	.04	.64	-.02	.34	.17	-.09	-.01			
Affect	.18	.37	.14	.19	.36	.42	.24	.03	.00	-.11	.04	-.09	(.12)		
Activity	.38	.28	.24	.34	.28	.38	.37	.19	.33	.14	.33	-.06	(.20)	(.21)	

TABLE 24

MULTITRAIT-MULTIMETHOD MATRIX
FOR OCCASION 26, SECOND ANALYSIS

	Self Ratings			Peer Ratings			Obs. Ratings			Test (ICL)			Event Records		
	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.	Inf.	Aff.	Act.
Self Ratings															
Influence															
Affect	(.48)														
Activity	(.38)	(.44)													
Peer Ratings															
Influence	.56	.34	.37												
Affect	.24	.28	.35	(.42)											
Activity	.35	.48	.37	(.51)	(.39)										
Obs. Ratings															
Influence	.43	.09	.13	.37	.03	.33									
Affect	.00	.04	.31	.17	.00	.29	(.34)								
Activity	.30	-.07	.28	.34	.17	.18	(.58)	(.45)							
Test (ICL)															
Influence	-.08	.18	.07	.02	-.10	-.03	-.01	.01	-.04						
Affect	-.19	-.01	.16	-.06	.09	.12	-.16	.07	-.03	(-.10)					
Activity	.15	.12	.28	.13	.20	-.03	-.10	.03	-.08	(.33)	(.07)				
Event Records															
Influence	.39	.00	.12	.36	-.12	.18	.54	.07	.39	.04	-.13	-.07			
Affect	.01	.07	.19	-.02	.04	.24	.08	.32	.12	-.10	.05	-.02	(.18)		
Activity	.19	.09	.19	.30	.05	.28	.15	.34	.36	.13	.39	.05	(.29)	(.18)	

the previous analysis, which were statistically significant and appeared to show moderately high convergences among the rating methods, show insignificant correlations when the activity-influence overlap is removed.

A case in point is found in Segment 1 (Tables 9 and 17). In Table 9, which includes the artifact, self ratings and peer ratings correlate .48, a significant and moderately strong convergence. But in Table 17, which has eliminated the artifact, the corresponding correlation is .19, which is not significantly different from zero. Numerous other cases in this and other pairs of matrices show the same pattern.

In this second analysis, there is still some convergence between self and peer ratings for influence and activity. Observer ratings and event records converge strongly for influence, but not activity. The validity of self and peer ratings is still blemished, however, by high method variance (high heterotrait-monomethod values). The observer ratings discriminate somewhat better and the event records discriminate very well.

It is worth noting that an unusual operational definition of activity had to be invoked in order to eliminate the artifactual relationship between influence and activity. Activity in an interpersonal situation is frequently defined in terms of talk time or communication rate. But here, activity came to mean bodily movements only: writing, change of posture, gestures, fiddling with pens, nervous twitches, etc. Thus, while this second multitrait-multimethod analysis exposes the lack of convergence among measurement methods more clearly than the first analysis, its definition of the activity trait is less closely related to previous research.

The artifact could have been removed by eliminating talking from influence, rather than from the activity scores. This would have left activity conceptually sound--talking and other motor actions. But it would have defined influence as not including effects of verbal communication--again, conceptually quite different from the usual meaning of the influence construct.

In sum, there is some convergent validity in the measurement methods used. Specifically, self and peer ratings converge for influence and activity, but not for affect. This convergence is stronger for influence than activity. Observer ratings converge with self and peer ratings on influence, but not on affect or activity. Observer ratings converge with event records when measuring influence. No other methods converge with event records on any variable. No methods converge with the Interpersonal Adjective Checklist at any point.

However, the rating methodologies are plagued by method variance. In general, they fail to discriminate between traits, although the discrimination between traits is better for observer ratings than for self or peer ratings. Lack of discriminant validity is further suggested by the presence of many relatively high heterotrait-heteromethod correlations.

The event records discriminate quite well between traits. This discrimination is better for affect than for influence and activity. They also show statistically significant correlations for convergences for influence, and in some cases for activity, with observer ratings. The heterotrait-heteromethod correlations for these two methods on these two traits tend to be low, relative to heterotrait-monomethod and heteromethod-monotrait correlations.

Changes Over Time

Behaviors

The foregoing section of this chapter exposed the widespread method variance and frequent lack of convergent validity which plague some common measurement methods in small group research. In spite of these measurement problems, however, the data seem worth examining for effects occurring over time. The ratings and event records, which are associated with specific time intervals during the research, provide an opportunity to evaluate temporal effects on subjects' behaviors.

The mean scores (from ratings and event records) for each observation period were averaged to provide an hour-by-hour score for each trait-method combination. The mean scores on influence and their standard errors for the normal-load hours are presented in Table 25.

TABLE 25

MEAN INFLUENCE SCORES FOR NORMAL-LOAD HOURS

Method \ Week	I		II		III		IV	
	M	est.σ_M[a]	M	est.σ_M[a]	M	est.σ_M[a]	M	est.σ_M[a]
Self Ratings	4.88	.81	4.91	.77	4.77	.83	4.85	.79
Peer Ratings	4.89	.67	4.85	.69	4.90	.55	4.80	.61
Observer Ratings	4.55	.84	4.29	1.08	4.67	.96	4.53	.98
Event Records	1.63	.69	1.51	1.09	1.84	.83	1.54	.82

[a]The reported means were computed as the average of the three mean scores of the hour. The appropriate index of variation is the standard error of the mean (est.σ_M).

A one-way Analysis of Variance (ANOVA) was performed on each variable, none of them yielding a significant F-ratio for influence. All statistical comparisons among these means failed to find significant differences.

A similar presentation of affect scores is contained in Table 26. Concerning affect, ANOVA results show a significant change over time for observer ratings. Specifically, the final week is significantly higher ($p \leq .01$) than any (and all) other weeks. None of the other measurement methods showed a significant change over time. The self ratings and peer ratings parallel the rise of the observer ratings during the final week, although not sharply enough to be statistically significant.

TABLE 26

MEAN AFFECT SCORES FOR NORMAL-LOAD HOURS

Method \ Week	I		II		III		IV	
	M	est.σ_M[a]	M	est.σ_M[a]	M	est.σ_M[a]	M	est.σ_M[a]
Self Ratings	5.00	.80	5.12	.66	5.07	.77	5.19	.83
Peer Ratings	5.12	.48	5.10	.41	5.14	.55	5.20	.65
Observer Ratings	4.52	.49	4.42	.47	4.50	.51	4.96	.84
Event Records	0.19	.27	0.11	.13	0.15	.19	0.15	.23

[a]The reported means were computed as the average of the three mean scores of the hour. The appropriate index of variation is the standard error of the mean (est.σ_M).

The means over time for activity scores and their standard errors are presented in Table 27. ANOVA's for self ratings, peer ratings, and event records all show significant changes over time in activity level.

Specifically, the self ratings show a significant ($p \leq .05$) rise between weeks II and III. This is parallelled by a significant ($p \leq .02$) rise in peer ratings at the same point. The rise in event records between the second and third weeks is significant at the .01 level. Only the observer ratings, which present a different pattern in mean values over time, fail to show a significant rise between weeks II and III.

TABLE 27

MEAN ACTIVITY SCORES FOR NORMAL-LOAD HOURS

Method \ Week	I		II		III		IV	
	M	est.σ_M[a]	M	est.σ_M[a]	M	est.σ_M[a]	M	est.σ_M[a]
Self Ratings	4.25	.82	4.30	.80	4.46	.81	4.53	.82
Peer Ratings	4.18	.64	4.22	.68	4.46	.63	4.44	.67
Observer Ratings	4.30	.84	3.97	1.01	4.27	.99	4.24	.96
Event Records	1.17	.41	1.12	.40	1.35	.43	1.27	.48

[a]The reported means were computed as the average of the three mean scores of the hour. The appropriate index of variation is the standard error of the mean (est.σ_M).

Thus, there are differences through time in the behaviors of small groups. Affect seems to rise during the fourth session, and activity clearly rises between the second and third sessions. In the case of affect, the observer ratings show a significant rise which is parallelled by a similar, though non-significant, rise in self and peer ratings. In the case of activity, three of the four methods show a significant rise between the second and third weeks.

Performance

In addition to changes in behavior, groups may change in their output characteristics over time. The products generated from the normal-load hours were analyzed by Analysis of Variance to assess the effects of time upon group output. The three task types were differentially affected by time, so the data for each task type is presented separately. The relations of the product dimensions to time for production, discussion, and problem-solving tasks are presented, respectively, in Tables 28, 29, and 30.

All three task types changed significantly in Action Orientation over time. Inspection of the means shows that production tasks were lower in Action Orientation during the second and third weeks than during the first or last. Discussion tasks rose steadily in Action Orientation over the four sessions. Problem-Solving tasks were higher in Action Orientation during the last two weeks than during the first two. So, while all task types differed in Action Orientation over time, they differed in different ways. Action Orientation is the only scale which showed significant time effects for problem-solving tasks.

Production and discussion groups changed significantly over time with respect to Originality and Involvement. On Originality, inspection of the means shows that production groups decreased over the first three sessions, but then increased during the fourth session. Discussion groups remained virtually constant in Originality for the first three weeks and increased during the fourth week.

Inspection of means shows that involvement increased sharply during Session II for production groups, then dropped to approximately

TABLE 28

MEANS SCORES FOR PRODUCT DIMENSIONS AND FOR
FOUR SESSIONS AND F-RATIOS BETWEEN SESSIONS (PRODUCTION GROUPS)

Dimensions	Means				F-ratio	p
	Week I	Week II	Week III	Week IV		
Action Orientation	2.67	1.89	2.00	2.78	2.89	$\leq$.05
Length	4.38	4.86	4.50	4.91	.24	ns
Originality	5.72	4.35	3.12	4.41	4.68	$\leq$.01
Outlook	3.78	4.12	4.02	3.95	1.18	ns
Quality of Presentation	4.20	5.70	5.12	4.89	3.68	$\leq$.025
Involvement	1.96	2.98	2.09	1.94	3.19	$\leq$.05
Adequacy	4.74	5.21	4.81	4.98	.25	ns
Creativity	4.63	4.57	3.64	4.48	1.02	ns

TABLE 29

MEANS SCORES FOR PRODUCT DIMENSIONS AND FOR
FOUR SESSIONS AND F-RATIOS BETWEEN SESSIONS (DISCUSSION GROUPS)

Dimensions	Means				F-ratio	p
	Week I	Week II	Week III	Week IV		
Action Orientation	2.56	2.94	4.04	4.33	9.74	≤.01
Length	2.54	2.52	3.13	2.70	1.02	ns
Originality	2.33	2.34	2.34	2.98	2.95	≤.05
Outlook	3.53	5.01	3.64	3.27	8.88	≤.01
Quality of Presentation	4.05	4.73	5.25	4.84	2.22	ns
Involvement	4.30	4.72	5.10	5.46	3.21	≤.05
Adequacy	3.92	4.82	4.41	4.64	1.33	ns
Creativity	2.24	2.72	2.76	3.40	4.13	≤.01

TABLE 30

MEANS SCORES FOR PRODUCT DIMENSIONS AND FOR FOUR SESSIONS AND F-RATIOS BETWEEN SESSIONS (PROBLEM-SOLVING GROUPS)

Dimensions	Means				F-ratio	p
	Week I	Week II	Week III	Week IV		
Action Orientation	5.86	5.61	6.52	6.43	4.79	≤.01
Length	2.62	2.95	3.36	2.68	1.68	ns
Originality	2.37	2.26	2.35	2.53	.79	ns
Outlook	4.29	4.42	4.27	4.29	.55	ns
Quality of Presentation	4.50	4.48	4.79	4.13	1.67	ns
Involvement	3.79	3.87	3.28	3.56	2.37	ns
Adequacy	4.88	5.18	4.99	4.59	1.11	ns
Creativity	2.45	2.64	2.84	2.81	.81	ns

its initial level during Sessions III and IV. For discussion groups, however, Involvement increased steadily over the four weeks.

Discussion groups alone were significantly affected on Outlook and Creativity. Inspection shows that outlook was considerably more positive during the second week than at any other session. After the second week it dropped to, and below, its initial level. Creativity increased over all four sessions for discussion groups.

In summary, all product dimensions are affected by time except Length and Adequacy. But only Action Orientation is affected in all three task types. When a dimension changes over time in two or three task types, it changes in a different way for each. There are changes on group performance through time, but the changes appear to be complex.

Changes Under Stress

Behaviors

Each weekly experimental session contained an hour of normal-load activity and an hour of stressing work load. During the first and third weeks the stressing load was an overload. During the second and fourth weeks it was an underload.

Changes in behaviors from one working condition to another can be examined on the basis of the ratings and event records. Table 31 presents, in parallel columns, the means and standard errors for normal and stress conditions on influence.

Examination of Table 31 reveals a strong tendency for influence behavior to rise under stress conditions. Seven of the comparisons between normal and stress conditions are significant, with the stress-

TABLE 31

COMPARISON OF INFLUENCE SCORES
BETWEEN NORMAL AND STRESS CONDITIONS

Method	Week	Normal Condition		Stress Condition		t
		M	est.σ_M[a]	M	est.σ_M[a]	
Self Ratings						
	I	4.88	.81	5.30	.73	3.00[c]
	II	4.91	.77	5.00	.99	-
	III	4.77	.83	5.14	.76	2.66[c]
	IV	4.85	.79	4.97	.72	-
Peer Ratings						
	I	4.89	.67	5.29	.54	3.64[c]
	II	4.85	.69	5.02	.65	-
	III	4.90	.55	5.15	.59	2.50[b]
	IV	4.80	.61	5.04	.57	2.27[b]
Observer Ratings						
	I	4.55	.84	5.15	.74	4.29[c]
	II	4.29	1.08	4.61	1.02	-
	III	4.67	.96	4.58	.87	-
	IV	4.53	.98	4.67	.91	-
Event Records						
	I	1.63	.69	1.71	.65	-
	II	1.51	1.09	1.67	.71	-
	III	1.84	.83	1.61	.65	-
	IV	1.54	.82	1.85	.88	2.07[b]

[a] The reported means were computed as the average of the three mean scores of the hour. The appropriate index of variation is the standard error of the mean (est.σ_M).

[b] $p \leq .05$

[c] $p \leq .01$

hour score higher than the normal hour score in every case. Of the remaining nine (insignificant) comparisons, seven means are higher under stress than under normal conditions.

While the tendency for influence to rise during the stress hours is general, the significant comparisons seem to cluster in the first and last weeks. Three methods (self, peer, and observer ratings) indicate a significant rise during Week I. Peer ratings and event records show a significant increase in influence during Week IV.

No significant change in influence is indicated during Week II, although all four methods show a tendency to rise between normal and stress hours. While both self and peer ratings agree on a significant rise during Week III, the observer ratings and event records fail to show any rise at that point. In fact, Week III showed a decrease (although insignificant) in influence as measured by observer ratings and event records.

Table 32 presents comparisons between normal and stress hours on affect behaviors. There is virtually no evidence of change in affect under stress. The only significant pair of means is the decrease in affect during Week I as measured by event records. While this lone significant comparison shows a decrease, the general trend in Table 32 is for affect scores to rise slightly during the stress hours. The failure to find significant shifts in affect under stress, along with the failure of all measurement methods to converge when measuring affect, are discussed in Chapter IV.

There is considerable evidence of rise in activity level as subjects encountered the stressing work load. Table 33 presents normal <u>vs</u>. stress comparisons in activity level. Self, peer, and

TABLE 32

COMPARISON OF INFLUENCE SCORES
BETWEEN NORMAL AND STRESS CONDITIONS

Method	Week	Normal Condition		Stress Condition		t
		M	est.σ_M[a]	M	est.σ_M[a]	
Self Ratings						
	I	5.00	.80	5.27	84	-
	II	5.12	.66	5.24	.75	-
	III	5.07	.77	5.23	.79	-
	IV	5.19	.83	5.39	.78	-
Peer Ratings						
	I	5.12	.48	5.29	.55	-
	II	5.10	.41	5.23	.52	-
	III	5.14	.55	5.22	.53	-
	IV	5.20	.65	5.36	.61	-
Observer Ratings						
	I	4.52	.49	4.68	.61	-
	II	4.42	.47	4.34	.49	-
	III	4.50	.51	4.64	.62	-
	IV	4.96	.84	5.09	.70	-
Event Records						[b]
	I	.19	.17	.51	.20	
	II	.13	.13	.23	.15	
	III	.15	.15	.16	.16	
	IV	.15	.23	.12	.23	-

[a] The reported means were computed as the average of the three mean scores of the hour. The appropriate index of variation is the standard error of the mean (est.σ_M).

[b] $p \leq .01$

TABLE 33

COMPARISON OF ACTIVITY SCORES
BETWEEN NORMAL AND STRESS CONDITIONS

Method	Week	Normal Condition		Stress Condition		t
		M	est.σ_M[a]	M	est.σ_M[a]	
Self Ratings						
	I	4.25	.82	4.68	.83	2.89[c]
	II	4.30	.80	4.59	.81	1.99[b]
	III	4.46	.81	4.75	.77	2.03[b]
	IV	4.53	.82	4.53	.77	-
Peer Ratings						
	I	4.18	.64	4.60	.57	3.84[c]
	II	4.22	.68	4.51	.63	2.44[b]
	III	4.46	.63	4.74	.56	2.60[b]
	IV	4.44	.67	4.47	.66	-
Observer Ratings						
	I	4.30	.84	4.83	.79	3.59[c]
	II	3.97	1.01	4.38	.91	3.33[c]
	III	4.27	.99	4.63	.78	2.23[b]
	IV	4.24	.96	4.53	.97	-
Event Records						
	I	1.17	.41	[illegible]	[illegible]	[illegible][b]
	II	1.12	.40	1.19	.39	
	III	1.35	.45	[illegible]	.39	
	IV	1.27	.48	1.27	.40	-

[a] The reported means were computed as the average of the three mean scores of the hour. The appropriate index of variation is the standard error of the mean (est.σ_M).

[b] $p \leq .05$

[c] $p \leq .01$

observer ratings all show significant increases in activity between normal and stress hours for Weeks I, II, and III. Event Records show a significant rise for Week I. The event records also show a rise, although non-significant, for Weeks II and III. Two of the four methods, peer ratings and observer ratings, show a rising trend during Week IV. In no case is there a decrease in activity level indicated as subjects moved from normal to stressing conditions. The fact that activity increases for underload stress as well as for overload stress may be worth noting.

Performance

Analyses of the written products show the effects of overload and underload working conditions on group performance. When the results of all groups are combined, every product dimension shows significant effects of the working conditions. Table 34 presents the relations between product dimensions and working conditions for all groups.

However, the working conditions had different effects on each task type. These differences are seen when the groups are divided by task type for analysis of stress effects on product dimensions. Tables 35, 36, and 37 present these effects for production, discussion, and problem-solving groups respectively.

From these tables it can be seen that the working conditions affected every task type on Length, Outlook, and Adequacy. Production and discussion groups were affected on Action Orientation. Production and problem-solving groups were affected on Quality of Presentation.

TABLE 34

MEANS SCORES FOR PRODUCT DIMENSIONS FOR EACH WORK LOAD CONDITION AND F-RATIOS BETWEEN CONDITIONS (ALL GROUPS)

Dimensions	Means			F-ratio	p
	Normal Load	Overload	Underload		
Action Orientation	4.37	4.27	4.93	638.69	≤.01
Length	3.25	2.71	4.70	91.23	≤.01
Originality	2.82	2.71	4.06	204.25	≤.01
Outlook	4.11	3.83	4.55	23.83	≤.01
Quality of Presentation	4.68	4.55	5.04	41.83	≤.01
Involvement	3.74	3.43	4.29	239.38	≤.01
Adequacy	4.74	4.59	5.60	187.49	≤.01
Creativity	3.06	2.82	3.90	112.34	≤.01

TABLE 35

MEANS SCORES FOR PRODUCT DIMENSIONS FOR EACH WORK LOAD CONDITION AND F-RATIOS BETWEEN CONDITIONS (PRODUCTION GROUPS)

Dimensions	Means			F-ratio	p
	Normal Load	Overload	Underload		
Action Orientation	2.34	2.31	2.35	20.29	≤.01
Length	4.71	3.79	5.18	9.06	≤.01
Originality	4.32	4.30	4.14	.08	ns
Outlook	3.97	3.47	4.22	8.95	≤.01
Quality of Presentation	5.01	5.07	5.78	3.94	≤.025
Involvement	2.23	2.55	2.84	1.70	ns
Adequacy	4.94	5.03	5.69	4.02	≤.025
Creativity	4.31	3.89	4.55	2.26	ns

TABLE 36

MEANS SCORES FOR PRODUCT DIMENSIONS FOR EACH WORK LOAD CONDITION AND F-RATIOS BETWEEN CONDITIONS (DISCUSSION GROUPS)

Dimensions	Means			F-ratio	p
	Normal Load	Overload	Underload		
Action Orientation	3.43	3.42	[illegible]	5.97	≤.01
Length	2.72	2.51	[illegible]	14.06	≤.01
Originality	2.47	2.25	2.73	4.81	≤.01
Outlook	3.93	3.97	4.92	4.24	≤.025
Quality of Presentation	4.73	4.83	5.35	1.86	ns
Involvement	4.88	5.04	6.01	7.57	≤.01
Adequacy	4.39	4.18	5.53	8.00	≤.01
Creativity	2.71	2.55	3.75	10.39	≤.01

TABLE 37

MEANS SCORES FOR PRODUCT DIMENSIONS FOR EACH WORK LOAD CONDITION AND F-RATIOS BETWEEN CONDITIONS (PROBLEM-SOLVING GROUPS)

Dimensions	Means			F-ratio	p
	Normal Load	Overload	Underload		
Action Orientation	6.13	6.09	6.26	.378	n.s.
Length	2.91	2.22	1.76	45.01	≤.01
Originality	2.32	2.10	1.54	6.68	≤.01
Outlook	4.32	3.93	4.53	6.80	≤.01
Quality of Presentation	4.47	4.02	4.29	4.48	≤.01
Involvement	3.62	3.13	4.11	12.03	≤.01
Adequacy	4.90	4.65	5.58	6.75	≤.01
Creativity	2.70	2.38	3.54	14.77	≤.01

And discussion and problem-solving groups were affected on Originality, Involvement, and Creativity. So, not only are there effects of stressful working conditions on performance in general, there are different effects of stress in each task type. These effects are discussed in Chapter IV.

Summary

The large amount of results presented in this chapter may be summarized as follows. Internal examination of the various measurement methods employed showed them to be acceptably consistent, reliable techniques. In the case of the Interpersonal Adjective Checklist, the present data conformed very well to the circumplex model upon which the test was built. In the case of the products, comparability to Hackman's earlier findings was outstanding.

Multitrait-multimethod analyses of convergent and discriminant validity revealed a large reservoir of method variance in the rating techniques. The ratings generally failed to discriminate well between traits. At some points there was convergent validity, with at least modest discrimination. In spite of its elegant internal features, the Interpersonal Adjective Checklist failed to converge with any other measurement method on any trait.

There were effects of time and effects of stress on group behaviors. Time and stress also affected group products. In the case of the products, time and stress had differential effects upon groups of different task types.

The results presented in this chapter are discussed in Chapter IV. There is an emphasis in Chapter IV on results which would have been

overlooked or misinterpreted in a less systematic study. The further discussion in Chapter V places emphasis upon the usefulness and potential of a systematic research strategy, such as the one which generated the present body of data.

CHAPTER IV: WILL THE REAL RELATIONSHIP PLEASE STAND UP?-- SOME EXERCISES IN INFERENCE FROM LIMITED EVIDENCE

This research was planned and executed as a methodological study. As such, the design was to pursue a research strategy which seemed promising by comparison to more frequently-used strategies. No hypotheses about the findings were set forth in advance and no specific substantive findings were sought. The analyses performed and reported in Chapter III were intended to suggest the comprehensiveness of the research strategy.

Two related questions determined the form of the present chapter, which discussed the results obtained: (1) What kinds of findings, ordinarily overlooked, are made available by this research strategy? and (2) What kinds of findings, frequently misinterpreted in other research, are elucidated by the present strategy? This chapter takes the form of citing selected findings which might have been obtained and overlooked or misinterpreted in other strategies of research. Then, the findings are shown, appropriately interpreted, in the context of the present systematic research.

Relationship Between Influence and Affect: The Method Variance Trap

Suppose that Mr. Brown, social psychologist, hypothesized a relationship between influence and affect in small groups. Specifically, he hypothesized that as group members become more influential, they are better liked. In order to test his hypothesis he studied several small

groups, measuring influence by self-ratings made by each subject and affect by peer ratings of how likeable each subject was.

On the basis of Table 9, which would include his data from a single-occasion study, the hypothesized relationship (.40) is significant ($p \leq .01$). He would have concluded that his hypothesis was confirmed.

Now, suppose that a second researcher accepts the relationship between influence and attraction, but hypothesizes that the relationship breaks down over time. He performs a two-occasion study on several groups, using the same measurement methods as Mr. Brown. Values in Tables 9 and 11 contain the second researcher's potential findings. The significant (.40) correlation obtained between influence and affect drops to an insignificant level (-.03) on the second occasion. He concludes that his hypothesis is confirmed.

As a further supposition, consider a third researcher who believed Mr. Brown's findings about the significant relationship. He hypothesized that the findings obtained by the second researcher may have been due to different levels of stress introduced in his two experimental sessions. In his experiment this third researcher used self ratings of influence, as did his predecessors. But for a more objective measure of affect, he employed a system comparable to the event records in the present study.

Tables 9 and 10 contain the third researcher's potential findings. Under normal working conditions (Table 9) he finds a significant ($p \leq .05$) correlation of .30 between influence and affect. And during the stressing overload condition (Table 10) the correlation drops to an insignificant .19. As a result, he concludes that stress, not time, was the likely cause for the deteriorating relationship between influence and affect in small groups.

As a finale to this series of experiments, a fourth researcher decides to follow up previous studies, measuring the relationship between influence and affect. However, he finds it convenient to modify the measurement methods somewhat. He allows each person to rate his own influence in the group, as before. But an impartial observer, watching the group, rates each member on affect.

Tables 9, 10 and 11 contain the fourth man's potential findings. On the first, normal-load (Table 9) occasion he fails to find a significant relationship ($r=.17$). However, the next hour's stressing load (Table 10) generates a significant ($p \leq .05$) relationship of .27. A week later (Table 11) he finds that the relationship is again insignificant ($r=.11$). His findings contradict all previous findings of a significant relationship between influence and affect in small groups under normal working conditions. And, his significant findings during the overload hour reverses another of the previous findings.

What can be said to clarify this tangled series of research results? In the context of the present design, all correlations cited are heterotrait-heteromethod values. These values have neither trait nor method in common and, therefore, should be low. The fact is that all rating methods used are plagued by method variance in both normal and stress conditions. And the convergent validity of all methods when measuring affect was extremely low. The findings obtained depend primarily upon the methods used, not upon the relationship supposedly measured. All four of the hypothetical researchers mentioned above capitalized on method variance and on chance.

Creativity in Groups: Confounding Task Types

Quite a separate stream of research is initiated, let us suppose, by Mr. Smith. His interest is in the outcome or performance of small groups on intellective tasks. He hypothesizes that group performance will improve over time on such tasks. By "improve" he means "become more creative."

Mr. Smith selects a set of intellective tasks, comparable in difficulty, and submits them to several groups of subjects. The groups meet for four consecutive weekly sessions to write responses to these tasks. Mr. Smith hires and trains "blind" judges to rate these products for creativity (by operations comparable to those used in the present study).

Table 29 contains a set of possible outcomes of Mr. Smith's research. The creativity scores rise from week to week, showing an overall significant ($p \leq .01$) gain in creativity in four weeks. On this basis, Mr. Smith concludes that his hypothesis is confirmed. (He may even be tempted by the opportunity to earn a sizeable supplementary income by "training" groups to become more creative.)

At another university a second researcher, also interested in group performance, feels that the creativity gain over time should be enhanced by tasks which stimulate creativity. However, he uses a different type of task than Mr. Smith, and redefines "goodness" as "originality," (in a manner similar to the originality ratings of the present study). He obtains quite different results. Table 28 contains potential results from this second researcher's study. It shows that originality does <u>change</u> significantly ($p \leq .01$) over time. But originality generally <u>declines</u>, rather than increases over time. In contrast to Mr. Smith's results, groups were significantly less

original after four weeks than they had been in the first week.

A third researcher is skeptical of both previous studies. He questions the operating conditions of the two studies, suspecting differential stress. To test his suspicions, he designs a study which includes normal load, overload, and underload conditions, using discussion tasks in small groups. As dependent measures, he checks both the originality and creativity of the products.

Table 36 contains the potential findings of the study outlined above. It shows significant changes in both originality and creativity under stress. Both tend to decrease with overload and increase with underload. The conclusion drawn is that stress, not time, is likely to account for changes in group products.

A fifth researcher, who doubts the "stress" explanation, tries to replicate some of the foregoing research. He reproduces the normal load, overload, and underload working conditions of the prior researchers. And he trains his product judges to rate for originality and creativity. His findings, potentially contained in Table 35, show no significant changes due to stress in either originality or creativity.

What has happened in this hypothetical stream of research? Does time affect group products? Does stress affect group products? Some answers are available from the present study.

The answers are "Yes, sometimes" for the effects of both time and stress on group products. Time does affect some types of group tasks on some rating dimensions, but not all types of tasks nor on all rating dimensions. Stress also affects products. But task types and rating dimensions are affected differentially.

What has happened in the hypothetical series of studies sketched above is that different researchers have unwittingly switched

task types on each other. They have also looked for slightly different qualities in the products. Their findings appear to be contradictory because their research was deceptively similar, while being different. Failing to note subtle but important distinctions between task types, they generalized a bit too far.

Patterning in Group Process: How Many Points Makes a Pattern?

A third hypothetical stream of research might be initiated by Mr. Jones, investigator of small group interaction patterns. His interest lies in the relationship between the giving and receiving of influence among group members. He hypothesizes a relationship between influence given and influence received.

Using an event records system such as the one designed in the present research, Mr. Jones observes several groups in a single-occasion study. His result, contained potentially in Table 38, is only a small (.16) insignificant correlation between influence given and influence received.

One of Mr. Jones' students pursues the question further, thinking that the relationship hypothesized may become apparent through time. He studies several groups, each of which has four weekly one-hour sessions. His results (the normal-load entries of Table 38) confirm his hypothesis. Indeed, the fourth session shows a significant ($p \leq .05$) positive correlation of .26 during the fourth week.

An industrial psychologist, upon reading of these results, considers their relevance for committees in his corporation. However, he suspects that working conditions may affect the relationship.

The industrial psychologist thus designs a study which varies the working conditions in a pattern comparable to that in the present

TABLE 38

CORRELATIONS BETWEEN INFLUENCE GIVEN AND INFLUENCE RECEIVED FOR SELECTED OCCASIONS

Week	Occasion	Condition	r	N	p
I	1	Normal Load	.16	62	ns
	5	Overload	.26	62	$\leq .05$
II	8	Normal Load	.18	62	ns
	12	Underload	.26	62	$\leq .05$
III	15	Normal Load	.12	62	ns
	19	Overload	.38	62	$\leq .01$
IV	22	Normal Load	.26	57	$\leq .05$
	26	Underload	.51	57	$\leq .01$

study (Figure 2). His potential results include all of Table 38. From these results he concludes that stress (both overload and underload) strengthens the relationship between influence given and influence received. He also notes that time strengthens the relationship in both stress and non-stress conditions.

In fact, the stream of research outlined above may have arrived at mistaken conclusions at several points. While the data drawn from the present study and contained in Table 38 are legitimate, they do not present the full picture. When several behavior samples are taken during each experimental hour, rather than just one, the picture becomes much more complex. The trends which seemed clear earlier then vanish. The full, segment-by-segment picture is contained in Table 39.

This Table shows great variability in the relationship between influence given and influence received. The relationship varies widely within each experimental hour, regardless of experimental conditions, until the final week. During Week IV the relationship is significantly positive throughout.

This chapter has aimed, in a somewhat unusual fashion, to demonstrate the usefulness of collecting a syetematic body of data on several traits, by multiple methods, through time. Many thoughtfully-conceived, carefully conducted studies run great risks of overlooking important findings or misinterpreting observed findings. In part, these risks exist because researchers use single methods of measurement, rather than multiple methods, in each study. Or, they measure only one trait. As an alternative, they may measure several traits by several methods, using one method per trait.

Other sources which contribute to the risk are the failure to conduct studies through time and under varied working conditions.

TABLE 39

CORRELATIONS BETWEEN INFLUENCE GIVEN AND INFLUENCE RECEIVED FOR ALL OCCASIONS

Week	Occasion[a]	Condition	r	N	p
	1	Normal Load	.16	62	ns
	2	Normal Load	.29	59	≤.05
	3	Normal Load	.13	45	ns
I	5	Overload	.26	62	≤.05
	6	Overload	.41	62	≤.01
	7	Overload	.04	42	ns
	8	Normal Load	.18	62	ns
	9	Normal Load	.18	62	ns
	10	Normal Load	.42	38	≤.01
II	12	Underload	.26	62	≤.05
	13	Underload	.18	62	ns
	14	Underload	.50	54	≤.01
	15	Normal Load	.12	62	ns
	16	Normal Load	.34	62	≤.01
	17	Normal Load	-.11	38	ns
III	19	Overload	.38	62	≤.01
	20	Overload	.35	62	≤.01
	21	Overload	.09	47	ns
	22	Normal Load	.26	57	≤.05
	23	Normal Load	.46	57	≤.01
	24	Normal Load	.40	44	≤.01
IV	26	Underload	.51	[illegible]	[illegible]
	27	Underload	.63	[illegible]	≤.01
	28	Underload	[illegible]	[illegible]	[illegible]

[a]The measurement occasions omitted were represented by too small a sample to yield stable correlations

In addition, and as was shown in the present study, differences in group tasks can easily lead different researchers to opposite conclusions.

The several hypothetical "research programs" used here for illustrative purposes are by no means unrepresentative designs. Indeed, they were chosen because they represent rather typical research approaches and substantive questions in the small group research area.

There are far more than a majority of single-relationship studies in the small group research field (McGrath and Altman, 1966). Furthermore, some sizeable bodies of "findings," which now have the status of "established principles," are based on relationships between two variables, each measured by the same, single method (most often, the ubiquitous self-report). The first illustration in this chapter shows how easily one can be tempted to over-interpret partial data, in directions that turn out to be systematic in error.

The general problem of disregard for task properties in studies of group performance has been discussed at length in a number of sources (e.g., Shaw, 1964; Hackman, 1966; McGrath and Altman, 1966). The more subtle but equally widespread problem of variations in "criterion" of group task performance effectiveness from one study to another has also been noted in the literature (e.g., Hackman, Jones and McGrath, 1967). The second "illustrative research program" described in this chapter exemplifies, from a body of systematic empirical data, how these task-related pitfalls can lead the investigator astray in the group performance research area.

Many theorists and researchers have emphasized the need for studies of patterning in group process through time. Very little work has been done on this topic. The third "illustrative research program" described here suggests some of the reasons why there has been such

limited effort in this area. Even the patterns of relationships between two quite similar variables (influence given and influence received) over time turn out to be very complex. Perhaps some important general relationships can be discovered by ignoring some of the fluctuations through time (as does Table 38). But this implies that somehow the data from time-points selected for inclusion are "true," and data from the omitted time-segments are "error." That is a very strong assumption for an investigator to make in an area which is so largely unexplored. Data presented in Table 39 give one pause regarding any straightforward inferences about the nature of the relationship between these two variables as a function of calendar time, task time, practice, or stress condition. It does make clear, though, that we cannot afford the "classic" one-occasion study if we are to learn about patterning of variables in group interaction.

There are a number of other "classical" research designs and problems from the small group research area which could have been selected for illustrative purposes here, and for which the present study contains a systematic set of relevant data. For example, this study has evidence (via several methods of measurement) on development and changes in "system balance"--regarding communication, attraction, and influence--in interpersonal systems. As another example, the study contains data which can be used to explore systematically the problem of phases in group development over time, again via several methods of measurement. These and other questions will be explored in later analyses.

But the main purpose here is in illustrating the potentialities that flow from the systematic approach to data collection which this study has followed--and, incidentally, to show that such a systematic strategy is feasible to carry out within relatively limited resources.

CHAPTER V: FURTHER DISCUSSION AND SUMMARY

Chapter IV, above, illustrated the usefulness of a systematic approach to research in small groups by exposing some types of errors which haunt other more restricted research strategies. The presentation of this final chapter is more positive, discussing some of the methodological implications of the findings of this research and pointing to potential gains available from systematic research approaches.

This chapter begins with some reflections on the measurement methods used in the present study. Second, there is a brief discussion of the traits (behaviors) which were measured. And third, some of the advantages of the systematic research are cited. These advantages are seen to accrue from both the multi-measure and multi-occasion features of the present research. The chapter closes with a brief summary.

Discussion of Specific Methods of Measurement

Interpersonal Adjective Checklist

The Interpersonal Adjective Checklist displayed satisfactory internal characteristics. Its circumplex ordering was virtually perfect. It discriminated clearly between the traits measured, including the "activity" trait for which it was not originally intended as a measure, and had reasonable reliability under the circumstances.

Yet, the Checklist failed to converge with any other measure on any of the traits. Its correlations with the other methods approach zero at every point. Even its correlations with the results of ratings

and event records which were obtained during the same hour as the administration of the Checklist were near-zero. Thus, although the ICL had fairly attractive internal features and the least method variance it was weakest in convergent validity.

Two comments stemming from this experience with the Checklist seem in order. First, it illustrates the severe constraints under which personality assessors must labor. Even the reliabilities, which by comparison to other available personality tests are considered good, place strict limits upon the measurement of personality traits. If the reliability of a test over a 2- to 4-month interval is .50 or .60, as it was for the Checklist in the present study, its validity can be no higher than .50 or .60. The fact seems to be that personality traits are very difficult to measure reliably, to say nothing of measuring validly. And if the multimethod convergent validity coefficients in this study are any indication, validity is a scarce commodity in much of our measurement.

The second comment emmanating from this experience with the Checklist is that personality traits may be situational. That is, "personality" may not consist of internal properties of the person which are invariant across time and place. Rather, it may consist of interactions of the person and the situation in which he finds himself. This implies that part of what is measured as "personality trait" actually "exists in" the measurement situation.

If so, then a reliability coefficient in the fifties or sixties across a 2- to 4-month time interval would seem quite encouraging. After all, the situation in which the testing took place was somewhat different on the two occasions and this would account for a moderate (as opposed to high) reliability. This explanation would be attractive

if the Checklist had correlated well with other measures taken during the same hour even though if it had correlated less well with measures taken at more distant times. This, unfortunately, was not the case. So, there is no evidence in this study that the Interpersonal Adjective Checklist was sensitive to the particular setting in which it was administered.

Ratings

All evidence in the present study points to an enormous method variance in all ratings methodologies, particularly self ratings and peer ratings. It would appear that an experimenter could find a significant positive relationship between virtually any pair of variables, provided that he measured both by a rating methodology. In the light of the observation of Webb, et. al. (1966) cited earlier, that ratings and questionnaires dominate social science as a measurement device, this finding casts serious doubt on a number of established relationships in the field which have been measured primarily by ratings and which thus become suspect.

In fairness to the observers, it must be pointed out that the observer ratings were plagued somewhat less by method variance than self or peer ratings. The observer ratings also showed some convergent validity with the event records, an achievement not attained by the other rating methods. Thus, data from the observer fared a bit better than data from subjects in regard to both discriminant and convergent validity.

Event Records

The event records system of observation designed in this research deserves further exploration and development. By comparison

to other observational systems (e.g., Bales, 1950b) it is a low-inference system. In addition, the inferences in this system are explicit, rather than hidden within the observer or the training process.

Potentially, the event records system may be exploited as a technique for measuring a wide variety of variables. Many of the behavioral indices reviewed by Argyle and Kendon (1967) and especially those of Sommer (1959, 1961, 1962, 1965) could be taken with great accuracy by this method.

Although the event records as taken in the present study involved the placement of a camera in the experimental room, operated directly by the observer, this need not have been the case. The methodology, in fact, is potentially quite unobtrusive. The camera could have been poised behind one-way glass outside the room. The event records methodology is also portable, capable of being carried into a wide variety of natural (field) settings. In many field settings it is possible to focus the camera on a group of subjects from a distance of twenty or thirty yards, using a telephoto lens. This makes it possible to collect data unobtrusively in a wide range of open settings. Thus, the event records methodology holds considerable potential as an unobtrusive, flexible, portable system for data-collection of behaviors in small groups.

Improvement of the event records system seems most likely to come from the use of videotape devices. The movie cameras used in the present study could have been replaced by videotape cameras. Videotaping provides a sound track which is perfectly synchronized with the visually-observed behaviors. (Currently-available videotape equipment is nearly as portable as a movie camera for use in field studies

and has no more lighting requirements than black and white films.)

A videotape apparatus was used as a training device for some observers in the present study. This brief experience demonstrated the feasibility of making videotapes of small group behaviors for subsequent analysis. It also demonstrated that, while the cost of videotape equipment is high, once it is acquired the cost of its use for data-collecting sessions is not prohibitively expensive, and use of videotape as a training device or mode for presentation of stimulus material is relatively economical.

Thus, there are features of merit in the event records system. Not all of the possibilities of this system were exploited in the present study, leaving room for considerable future development. One of the ways in which the event records system may develop is in the use of videotape.

Discussion of Specific "Traits" in Interpersonal Interaction

It is difficult to make general statements about the traits (behaviors) measured in the present research. The general method variance of the ratings has been noted. This indicates restricted discrimination between traits for rating methods. Also noted was the discrimination but non-convergence of traits for the Interpersonal Adjective Checklist. The event records, while discriminating moderately well, converged with another method (observer ratings) on only one trait (influence). These weaknesses of the measurement methods make it impossible to be confident of any conclusion concerning the traits as unitary constructs.

Under the original plan of analysis, there appeared to be convergence among several methods on both influence and activity.

However, the operational definitions of influence and activity, designed to reflect prevalent conceptions of these traits, connected them artifactually. When the artifact was removed, for ratings and for event records, the method variance of the event records and the convergence of the rating methods both decreased.

The remaining convergence in the multitrait-multimethod matrices stood between observer ratings and event records on influence, at a statistically significant but not overpowering level. For activity, convergence became very weak when its overlap with influence had been removed. And for affect, which never enjoyed anything but method variance, there was no convergence at any point.

One conclusion to be drawn from this series of analyses is that to remove the method variance from the measures is to change the concept being measured. The concepts available for measurement, while widely used, are so method-bound that results of their measurement are primarily a matter of method variance. In this case, method variance was decreased by re-defining "activity" in a non-interpersonal way. Another conclusion to be drawn from these analyses is that we must remain quite skeptical as to the convergent and discriminant validity of these traits, with the possible exception of influence, even though these three traits have been identified in the literature as pre-eminent variables in small group interaction.

Potential Gains From Use of a Systematic Strategy

Gains From a Multi-Measure Approach

The basic multitrait-multimethod argument (Campbell and Fiske, 1959; Webb, et. al., 1966) is that any measurement consists of a trait-method complex, which contains a "true" component and some

"error." Different methods of measuring behavior must, by definition, share the "true" component, but will tend to contain different "error." For any particular measurement method the errors are likely to pattern or contain systematic bias regardless of the "trait" being measured. The use of multiple measures which are subject to different patterns of error is proposed as a means of separating the "true" trait component from the method-linked error.

Even when a completely-crossed multitrait-multimethod design is not feasible, having the goal of multiple measures can be beneficial to the researcher. If an investigator adopts the goals of developing multiple measures for each of his concepts, he is likely to clarify the concepts (traits, behaviors) being measured considerably. Specifically, he denies himself the use of "traits" for which there is only one possible measure. When only one method is (or can be) used to measure a trait over a series of experiments, the trait-method complex becomes the unitary construct. The trait thus tends to reflect some of the error peculiar to the particular measurement method with which it is yoked.

It may be granted that there are some concepts which can be measured by only one method. Perhaps self reports are the only method which can asses certain variables dealing with psychological experience. If so, such variables are eternally method-bound. There is no way in which the error pattern peculiar to the measurement method can be separated from the trait always measured thereby. Such concepts are always method-bound, and hence limited in the generalizability of results obtained from their use.

Thus, while use of multiple methods may not always be feasible, they are always desirable. They provide a means of separating method (error) variance from trait ("true") variance. Besides, the

implementation of this approach in specific experiments can stimulate the investigator to develop methodologically and conceptually "clean" constructs.

Gains From a Multi-Occasion Approach

The gains from a multi-occasion approach are numerous. The most obvious is that changes in behaviors through time may be observed. The call for this type of knowledge, as noted in Chapter I, has come repeatedly from many sources.

Another kind of gain from the multi-occasion strategy deserves mention, however. Consider that traditional nomothetic research has tended to measure behaviors over a large number of subjects on a single occasion. In part, this tendency must be due to the convenience of single-occasion studies and to the availability of statistical power for this type of design. It has, in the process, tended to concentrate on traits which are common to large numbers of people, rather than traits which tend to identify individuality.

In contrast, the idiographic approach has pressed for study of the traits which identify individuality. In so doing, the idiographic approach has tended to conduct multi-occasion studies of single individuals. What has been lacking is the statistical power to handle many of the data collected idiographically.

Now, one can manipulate these concepts in an interesting way. For example, the traditional single-occasion, large population study could be viewed as an idiographic study of an occasion. With this view, the occasion is the object being studied and the people are a series of sample points. In contrast, the study of a single individual over time can be considered idiographic with respect to population, but

nomothetic with respect to occasion. The individual is the object being studied, and the occasions of measurement are the sample points.

It follows that if a multitrait-multimethod matrix can be constructed for an occasion, as has been done for each of several occasions in the present study, one could be constructed for an individual person. The matrix, with traits nested inside the methods, would take exactly the same form as Tables 9-24. Convergent and discriminant validity in the matrix would point to valid methods of measurement <u>for that individual</u> (over a series of time-ordered occasions).

It would seem useful to pursue such a strategy whenever one was interested in obtaining valid indices of behavior for a specific individual. It might also be a useful technique for identifying subsets of individuals for whom a certain measurement method yields valid results. Above all, this would appear to be a useful strategy for those areas of behavioral science which deal with very large social systems (organizations, total societies), for which only one or a few "cases" exist, but for which comprehensive measurements over time can be obtained. While data from the present study are not analyzed or discussed from the above perspective here, such analyses will be pursued in future explorations of these data.

Summary

This has been a methodological study, designed to show the value of a systematic comprehensive measurement approach to research on small groups. The main themes of the study have been (1) the value of using multiple measurement methods and (2) the value of measurement over a series of time-ordered occasions.

The results presented are a selected subset of the total array of results. Their selection was designed to show the power of a systematic body of data. These results have been discussed in two ways. The discussion in Chapter IV shows how some more limited research designs may lead to interpretive problems or errors which could be avoided. Chapter V looks at specific methods and traits studied and considers the advantages of the present approach to research on small groups.

LIST OF REFERENCES

Altman, I., & Haythorn, W. W. Interpersonal exchange in isolation. Sociometry, 1965, 28, 411-426. (Republished: Defense Documentation Center, AD662091.)

Appley, M. H., & Trumbull, R. Psychological stress. New York: Appleton-Century-Crofts, 1967.

Argyle, A., & Kendon, A. The experimental analysis of social performance. In L. Berkowitz, (Ed.), Advances in experimental social psychology. Vol. 3. New York: Academic Press, 1967, Pp. 55-99.

Bales, R. F. A set of categories for the analysis of small group interaction. American Sociological Review, 1950, 15, 257-263. (a)

Bales, R. F. Interaction process analysis: A new method for the study of small groups. Reading, Mass.: Addison-Wesley, 1950. (b)

Bales, R. F., & Gerbrands, H. The interaction recorder. Human Relations, 1948, 1, 456-463.

Bales, R. F., & Slater, P. E. Role differentiation in small decision-making groups. In T. Parsons & R. F. Bales (Eds.), Family, socialization, and interaction process. Glencoe, Ill.: Free Press, 1955, Pp. 259-306.

Bales, R. F., & Strodtbeck, F. L., Phases in group problem-solving. Journal of Abnormal Social Psychology, 1951, 46, 484-496. In W. Vinacke, R. Wilson, & M. Meredith (Eds.), Dimensions of social psychology. Chicago: Scott, Foresman, 1964, Pp. 415-425.

Borgatta, E. F., Cottrell, L. S., & Mann, J. M. The spectrum of individual interaction characteristics: An interdimensional analysis. Psychological Reports, 1958, 4, 279-319.

Campbell, D. T. Recommendations for APA test standards regarding construct trait or discriminant validity. American Psychologist, 1960, 15, 546-553.

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Carter, L. F. Some research on leadership in small groups. In H. Guetzkow (Ed.), Groups, leadership and men: Research in human relations. Pittsburgh: Carnegie Press, 1951, Pp. 146-157.

Cartwright, D., & Zander, A. (Eds.) Group dynamics. (2nd Ed.) New York: Row, Peterson, 1960.

Deutsch, M. An experimental study of the effects of cooperation and competition upon group process. Human Relations, 1949, 2, 199-231.

Dewey, J. How we think. Boston: Heath, 1910.

Dulany, D. E. Awareness, rules, and propositional control: A confrontation with S-R behavior theory. In T. Dixon & D. Horton, Verbal behavior and general behavior theory. New York: Prentice-Hall, 1968.

Faucett, R. E., & Newman, P. P. Operation hideout: Preliminary report. United States Navy Medical Research Laboratory Report, 12 (13, Whole No. 228), iii, New London, Conn., July, 1953.

Firth, R. Elements of social organization. Boston: Beacon, 1963.

Garner, W. R. Context effects and the validity of loudness scales. Journal of Experimental Psychology, 1954, 48, 218-224.

Garner, W. R., Hake, H. W., & Eriksen, C. W. Operationism and the concept of perception. Psychological Review, 1956, 63, 149-159.

Hackman, J. R. Effects of task characteristics on group products. Technical Report No. 5, Air Force Office of Scientific Research, 1966.

Hackman, J. R., Jones, L. E., & McGrath, J. E. A set of dimensions for describing the general properties of group-generated written passages. Psychological Bulletin, 1967, 6, 379-390.

Hase, H. D., & Goldberg, L. R. Comparative validity of different strategies of constructing personality inventory scales. Psychological Bulletin, 1967, 67, 231-248.

Haythorn, W. W., & Altman, I. Personality factors in isolated environments. In M. H. Appley and R. Trumbull (Eds.), Psychological stress. New York: Appleton-Century-Crofts, 1967, Pp. 363-386.

Haythorn, W. W., Altman, I., & Myers, T. J. Emotional symptomatology and subjective stress in isolated pairs of men. Journal of Experimental Psychology, 1966, 1, 290-305.

Heinicke, C., & Bales, R. F. Developmental trends in the study of small groups. Sociometry, 1953, 16, 7-38.

Hemphill, J. K., & Westie, C. M. The measurement of group dimensions. Journal of Psychology, 1950, 29, 325-342.

LaForge, R. Research use of the ICL. Oregon Research Institute Technical Report No. 4, Eugene, Oregon, 1963.

Leach, E. R. Political systems of highland Burma. London: London School of Economics and Political Science, 1954.

Leary, T. F. Interpersonal diagnosis of personality. New York: Ronald Press, 1957.

Levine, A. S. Habitability and motivation as related to the Polaris submarine. United States Navy BuPers, Personnel Research Division Report, FBM-11, December, 1958.

Lippitt, R. Training in community relations: A research exploration toward new group skills. New York: Harper, 1949.

McGrath, J. E. A descriptive model for the study of interpersonal relations in small groups. Journal of Psychological Studies, 1963, 14, 89-116.

McGrath, J. E. & Altman, I. Small group research: A synthesis and critique of the field. New York: Holt, Rinehart, & Winston, 1966.

Murphy, L. B. Social behavior and child personality: An exploratory study of some roots of sympathy. New York: Columbia Press, 1937.

Nadel, S. F. The theory of social structure. New York: Free Press, 1964.

Newcomb, T. M. The acquaintance process. New York: Holt, Rinehart, & Winston, 1961.

Orne, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. American Psychologist, 1962, 17, 776-783.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. The measurement of meaning. Urbana: University of Illinois Press, 1957.

Parten, M. Social play among preschool children. Journal of Abnormal and Social Psychology, 1933, 28, 136-147.

Ritch, T. C. Report on the effects of prolonged snorkelling on the health of the officers and men and on the general habitability of the Guppy-Snorkel submarine USS Trumpetfish (SS425), United States Navy Medical Research Laboratory Report No. 132, New London, Conn., April, 1948.

Schaefer, E. S. Converging conceptual models for maternal behavior and for child behavior. Paper presented at the Second Annual Conference of the Social Science Institute, Washington University, St. Louis, 1960.

Schutz, W. C. FIRO: A three dimensional theory of interpersonal behavior. New York: Holt, Rinehart, & Winston, 1958.

Schutz, W. C. The ego, FIRO theory and the leader as completer. In L. Petrullo & B. M. Bass (Eds.), Leadership and interpersonal behavior. New York: Holt, Rinehart, & Winston, 1961.

Scott, W. A. Social desirability and individual conceptions of the desirable. Journal of Abnormal and Social Psychology, 1963, 67, 574-585.

Shaw, M. E. Communication networks. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology. Vol. I. New York: Academic Press, 1964.

Sommer, R. Studies in personal space. Sociometry, 22, 1959, 247-260.

Sommer, R. Leadership and group geography, Sociometry, 24, 1961, 99-109.

Sommer, R. The distance for comfortable conversation, Sociometry, 1962, 25, 111-116.

Sommer, R. Further studies of small group ecology. Sociometry, 1965, 28, 337-348.

Smith, W. M. Observations over the lifetime of a small isolated group: Structure, danger, boredom, and vision. Psychological Reports, 1966, 19, 475-514.

Talland, G. A. Task and interaction process: Some characteristics of therapeutic group discussion. Journal of Abnormal and Social Psychology, 1955, 50, 105-109.

Theodorson, G. A. Elements in the progressive development of small groups. Social Forces, 1953, 31, 311-320.

Tuckman, B. W. Developmental sequence in small groups. Psychological Bulletin, 1965, 63, 384-399.

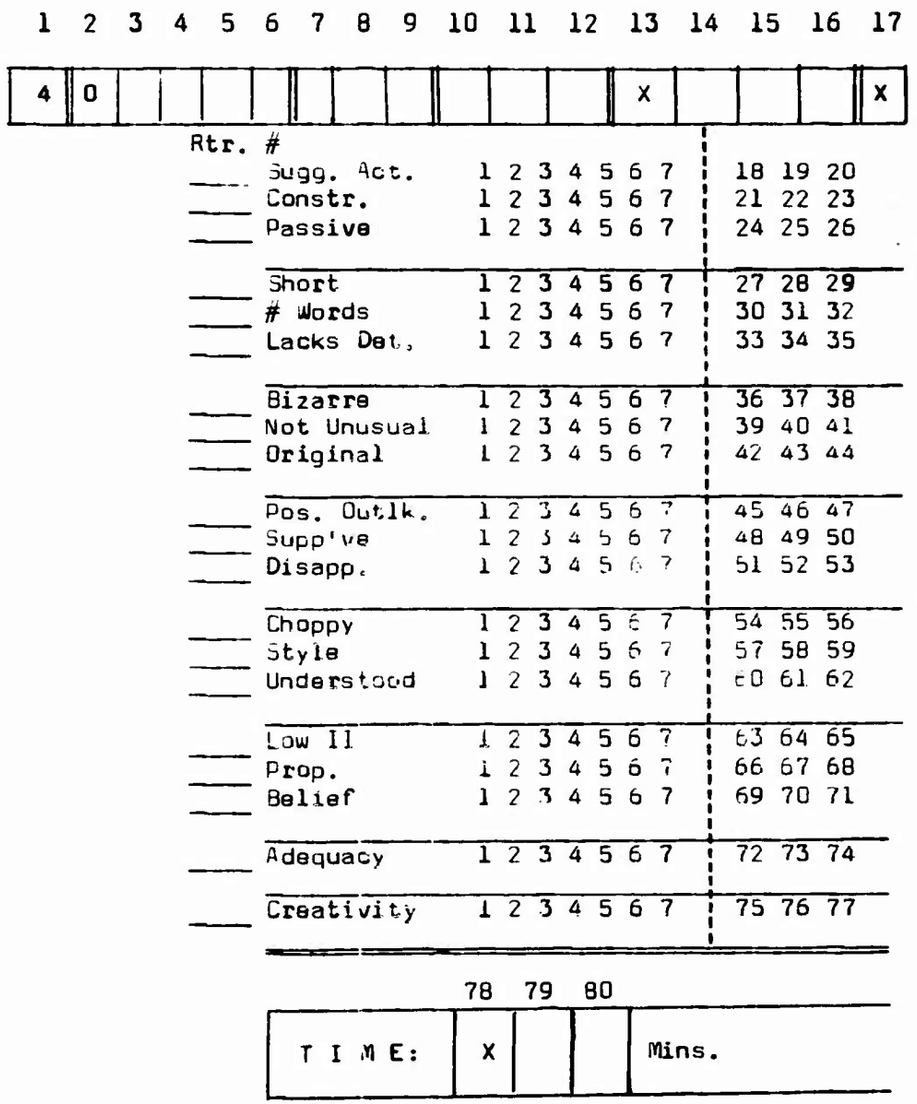
Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. Unobtrusive measures. Chicago: Rand McNally, 1966.

Weybrew, B. B. Psychological and Physiological effects of long periods of submergence. Analysis of data collected during a 265-hour completely submerged habitability cruise made by the USS Nautilus (SSN571), United States Naval Medical Research Laboratory Report No. 16 (3, Whole, No. 281), iv, New London, Conn., February, 1957.

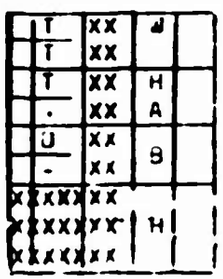
Weybrew, B. B. Psychological problems of prolonged marine submergences. In M. Burns, M. Chambers, & E. Hendler (Eds.), Unusual Environments and human behavior: Physiological and psychological problems of men in space. New York: Free Press, 1963, Pp. 87-125.

Weybrew, B. B. Patterns of psychophysiological response to military stress. In M. H. Appley & R. Trumbull (Eds.), Psychological stress. New York: Appleton-Century-Crofts, 1967.

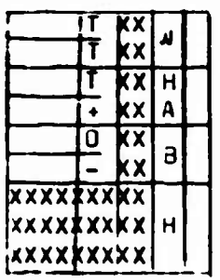
APPENDIX


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
4 0 X X
Rtr. #
Sugg. Act. 1 2 3 4 5 6 7 18 19 20
Constr. 1 2 3 4 5 6 7 21 22 23
Passive 1 2 3 4 5 6 7 24 25 26
Short 1 2 3 4 5 6 7 27 28 29
Words 1 2 3 4 5 6 7 30 31 32
Lacks Det. 1 2 3 4 5 6 7 33 34 35
Bizarre 1 2 3 4 5 6 7 36 37 38
Not Unusual 1 2 3 4 5 6 7 39 40 41
Original 1 2 3 4 5 6 7 42 43 44
Pos. Outlk. 1 2 3 4 5 6 7 45 46 47
Supp've 1 2 3 4 5 6 7 48 49 50
Disapp. 1 2 3 4 5 6 7 51 52 53
Choppy 1 2 3 4 5 6 7 54 55 56
Style 1 2 3 4 5 6 7 57 58 59
Understood 1 2 3 4 5 6 7 60 61 62
Low II 1 2 3 4 5 6 7 63 64 65
Prop. 1 2 3 4 5 6 7 66 67 68
Belief 1 2 3 4 5 6 7 69 70 71
Adequacy 1 2 3 4 5 6 7 72 73 74
Creativity 1 2 3 4 5 6 7 75 76 77
78 79 80
T I M E: X Mins.

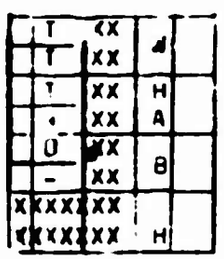
A



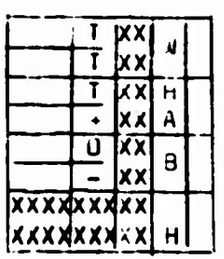
B



D



C



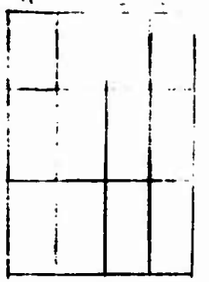
I N F L U E N C E

(Given)

Talking x--N-l--

Being looked at (# looking)

Reciprocal looking at person while talking.

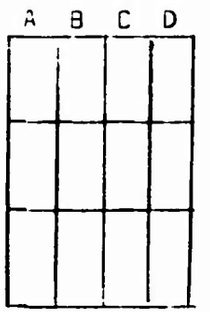


(Received)

Being talked to (# of talkers

Looking at (1 for each)

Reciprocal looking at speaker.



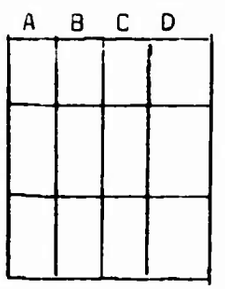
A F F E C T (Positive)

(Given)

Smile

Being Looked at (# of people)

Reciprocal looking while smiling

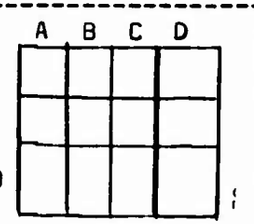


(Received)

Looking at smile

Smile

Reciprocal looking at smile



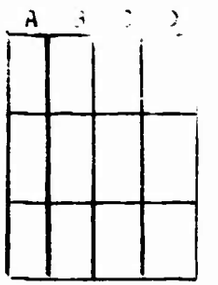
A F F E C T (Negative)

(Given)

Frown, etc.

Being looked at (# of people)

Reciprocal looking while frowning

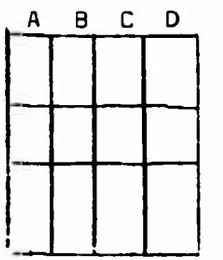


(Received)

Looking at frown

Frowns, etc.

Reciprocal looking at frown



A C T I V I T Y

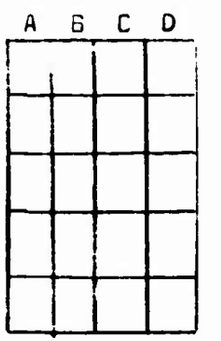
Talks (# of times)

Writes

Hands/Arms Move

Body moves

Head moves



Group # _____ Session I II III IV

Date,_____ 1st. Hour: Task 1 2 3 4

Observer _____ 2nd. Hour: segment 1 2 3

Room _____ Film rdg. _____

During the first 5 seconds of each minute, as signalled by the Hunter timers:

1) take three (3) photos, spaced evenly over the 5 seconds,
2) observe who is talking while you take the photos.

When the signal tells you the 5 seconds are over, record who was speaking by placing a check in the appropriate box. If more than one person speaks during the 5 seconds, record them in the order of speaking, (1, 2, 3)

Take one blank frame (hand over lens) between each task (1st. hour) or between each 15-minute segment (2nd hour) and note this blank frame. Take 3 blank frames (hand over lens) at the beginning of each hour.

One (or 3) blanks: one three (circle)

	A	B	C	D	No One
Photo 1,2,3					
Photo 4,5,6					
Photo 7,8,9					
Photo 10,11,12					
Photo 13,14,15					
Photo 16,17,18					
Photo 19,20,21					
Photo 22,23,24					
Photo 25,26,27					
Photo 28,29,30					
Photo 31,32,33					
Photo 34,35,36					
Photo 37,38,39					
Photo 40,41,42					
Photo 43,44,45					
2nd hour. Stop for ratings here.					
Photo 46,47,48					
Photo 49,50,51					

	A	B	C	D	No One
Photo 52,53,54					
Photo 55,56,57					
Photo 58,59,60					
Photo 61,62,63					
Photo 64,65,66					
Photo 67,68,69					
Photo 70,71,72					
Photo 73,74,75					
Photo 76,77,78					
Photo 79,80,81					
Photo 82,83,84					
Photo 85,86,87					
Photo 88,89,90					
Photo 91,92,93					
Photo 94,95,96					
Photo 97,98,99					
Photo 100-102					

NAME: _____

DIRECTIONS: This sheet contains a list of descriptive words and phrases which can be used to describe people. Mark those adjectives which are generally descriptive of _____.

1. often admired
2. respected by others
3. good leader
4. likes responsibility
5. self-confident
6. self-reliant and assertive
7. business like
8. likes to compete with others
9. hard-boiled when necessary
10. stern but fair
11. irritable
12. straightforward and direct
13. resents being bossed
14. skeptical
15. hard to impress
16. touchy and easily hurt
17. easily embarrassed
18. lacks self-confidence
19. easily led
20. modest
21. often helped by others
22. very respectful to authority
23. accepts advice readily
24. trusting and eager to please
25. always pleasant and agreeable
26. wants everyone to be like him
27. sociable and neighborly
28. warm
29. kind and reassuring
30. tender and soft-hearted
31. enjoys taking care of others
32. gives freely of self
33. tries to be too successful
34. expects everyone to admire him
35. manages others
36. dictatorial
37. somewhat snobbish
38. egotistical and conceited
39. selfish
40. cold and unfeeling
41. sarcastic
42. cruel and unkind
43. frequently angry
44. hard-hearted
45. resentful
46. rebels against everything
47. stubborn
48. distrusts everybody
49. timid
50. always ashamed of self
51. obeys too willingly
52. spineless
53. hardly ever talks back
54. clinging vine
55. likes to be taken care of
56. will believe anyone
57. wants everyone's love
58. agrees with everyone
59. friendly all the time
60. loves everyone
61. too lenient with others
62. tries to comfort everyone
63. too willing to give to others
64. spoils people with kindness

NAME: _____

55. well thought of
56. makes a good impression
67. able to give orders
68. forceful
69. self-respecting
70. independent
71. able to take care of self
72. can be indifferent to others
73 can be strict if necessary
74. firm but just
75. can be frank and honest
76. critical of others
77. can complain if necessary
78. often gloomy
79. able to doubt others
80. frequently disappointed
81. able to criticize self
82. apologetic
83. can be obedient
84. usually gives in
85. grateful
86. admires and imitates others
87. appreciative
88. very anxious to be approved of
89. cooperative
90. eager to get along with others
91. friendly
92. affectionate and understanding
93. considerate
94. encourages others
95. helpful
96. big-hearted and unselfish
97. always giving advice
98. acts important
99. bossy
100. dominating
101. boastful
102. proud and self-satisfied
103. thinks only of himself
104. shrewd and calculating
105. impatient with others' mistakes
106. self-seeking
107. outspoken
108. often unfriendly
109. bitter
110. complaining
111. jealous
112. slow to forgive a wrong
113. self-punishing
114. shy
115. passive and unaggressive
116. meek
117. dependent
118. wants to be led
119. lets others make decisions
120. easily fooled
121. too easily influenced by friends
122. will confide in anyone
123. fond of everyone
124. likes everybody
125. forgives anything
126. oversympathetic
127. generous to a fault
128. overprotective of others

FORM 2-g

OBSERVER POST-TASK RATING FORM (Group)

Task_____1, _____2

Order ___1, ___2, ___3, ___4

Group #_____ Date_____

Observer_____

SUBJECT:

A_____

B_____

C_____

D_____

Rate each of the four group members on each of the following scales. For each man (A,B,C,D) place his letter in the appropriate box. More than one man may appear in any given box. But each man should appear only once on each scale.

Functioned smoothly in gps. task performance	3 2 1 0 1 2 3	Obstructed groups' task performance
Talked constantly	7 6 5 4 3 2 1	Never talked
Moved hands or body constantly	7 6 5 4 3 2 1	Never moved hands or body
Interested in gps. activity (task relevant or not)	7 6 5 4 3 2 1	Bored with gps. activity
Tried to influence group	3 2 1 0 1 2 3	Submissive in group; accepted influence from others
Showed liking for others	3 2 1 0 1 2 3	Showed dislike for others
Seemed liked by others	3 2 1 0 1 2 3	Seemed disliked by others

NOW PLACE A CHECK SOMEWHERE ON THESE LINES TO INDICATE THE RATING YOU WOULD GIVE THE GROUP AS A WHOLE

Pooled resources smoothly in task performance	3 2 1 0 1 2 3	Confusion in task performance
Active pace: high tempo of group activity (task relevant or not)	7 6 5 4 3 2 1	Slowness as a group: low activity level

Form 2-g

DIRECTIONS: Note the roles played by each member of the group. Place a member's letter (A,B,C,D) in each space, according to the role(s) played. More than one letter is allowed in each space; any any member may be listed more than once.

Leader_____

Scribe/recorder_____

Inactive Member_____

Normal/Active Member_____

FORM S-2-g (Group)

SUBJECT POST-TASK RATINGS

The four members of your group are: (circle your own letter)

A _____

B _____

C _____

D _____

Evaluate yourself and EACH OTHER MEMBER of your group on the scales below. Place each man's letter in one of the boxes on each scale. More than one letter may appear on any one box. Consider ONLY the task you have just completed.

Functioned smoothly in gps. task performance	3 2 1 0 1 2 3	Obstructed groups' task performance
Talked constantly	7 6 5 4 3 2 1	Never talked
Moved hands or body constantly	7 6 5 4 3 2 1	Never moved hands or body
Interested in gps. activity (task relevant or not)	7 6 5 4 3 2 1	Bored with gps. activity
Tried to influence group	3 2 1 0 1 2 3	Submissive in group; accepted influence from others
Showed liking for others	3 2 1 0 1 2 3	Showed dislike for others
Seemed liked by others	3 2 1 0 1 2 3	Seemed disliked by others

NOW PLACE A CHECK SOMEWHERE ON THESE LINES TO INDICATE THE RATING YOU WOULD GIVE THE GROUP AS A WHOLE

Pooled resources smoothly in task Performance	3 2 1 0 1 2 3	Confusion in task performance
Active pace; high tempo of group activity (task relevant or not)	7 6 5 4 3 2 1	Slowness as a group; low activity level

Relationship to the Study of Rosemary Lowe

The present study is part of a larger program of research on psychological stress, with Dr. Joseph E. McGrath as principle investigator. The role of the present study within this larger program is to explore relevant methodological problems and to develop a low-inference observational method for the study of small groups under stress.

This study has a particularly close relationship to a study carried out simultaneously by Rosemary Lowe. As described in Chapter II, Ss for the present study were recruited from Lowe's subject population. Ss in the present study were compared against Lowe's total population on the two dimensions of the ICL (Dominance and Affect) and on activity level as measured by the number of items endorsed on the ICL. Results of these tests are presented in Table 40. No significant differences were found. From the results of the Interpersonal Adjective Checklist (ICL), it appears that a representative subset of Lowe's Ss was recruited.

TABLE 40

TEST OF DIFFERENCES BETWEEN
LOWE'S SUBJECTS AND STAPERT'S SUBJECTS
ON THE INTERPERSONAL ADJECTIVE CHECKLIST

Trait	Lowe's (N=369)		Stapert's (N=59)		t	p
	M	SD	M	SD		
Dominance	1.68	3.3	1.61	4.5	.11	ns
Affect	-.80	4.1	-2.40	7.6	1.57	ns
Activity	16.00	6.1	14.60	5.4	1.79	ns

Both studies used Hackman's tasks. The Lowe study used only twelve of the tasks, four of each task type, while the present study used all 108 tasks (36 of each type). The 12 tasks used by Lowe were used in the present study as the last two tasks of each of the two overload hours.

In addition to the common subject population and common tasks, there were other ties between the two studies. The same set of trained assistants acted as observers in the experiment. And the judges who rated written products were jointly trained and employed for rating products from both studies.

These connections between the two studies have provided relationships which allow comparisons between the studies. At the same time, it was possible to collect distinct bodies of data which could be analyzed separately and independently for each study.

Note on 3-man vs. 4-man Groups

Both 3-man and 4-man groups are included in the present study. The original intent was to study only 4-man groups. The 3-man groups were accepted only where it was impossible to schedule a 4-man group. In fact, all of the experimental groups were originally scheduled as 4-man groups. At least two attempts were made to have all four members present at the first session. If both efforts failed, the group was run as a 3-man group throughout the experiment.

Some 4-man groups had difficulty maintaining their full membership for all four experimental sessions. In spite of numerous telephoned reminders and other efforts to keep all Ss who began the experiment, some groups lost members. These experiences were some indication of why multiple-occasion research on groups is seldom done.

The numerical differences between 3-man and 4-man groups raised some problems for data-handling and statistical analysis. The scores which required re-scaling before analysis were peer ratings of all traits and event records of influence and affect. The peer ratings of a subject were always averaged before statistical analysis. Thus, in a 3-man group, the sum of the peer ratings required division by two and in 4-man groups required division by three. This was a simple adjustment to make.

Adjustment of the event records was a bit more complicated. Analysis of the scoring procedures showed that equivalent behaviors in 3- and 4-man groups required 3-man influence scores to be multiplied by

1.50 and the 3-man affect scores to be multiplied by 1.25. This made scores from 3-man and 4-man groups numerically comparable.

However, there is a more fundamental question involved than the re-scaling. That is, does group size play a part in eliciting different levels of influence, affect, or activity in the group? This question was answered, in part, by comparing mean scores from 3- and 4-man groups on each variable (each trait measured by each method). Table 41 presents these comparisons.

Of the 15 comparisons, 14 are not significantly different. The event records measure of affect shows a small (.05 unit) difference which is barely statistically significant ($p \leq .05$). This sole significant difference seems as likely to have been caused by a slight under-correction in the re-scaling procedures as by any real difference between the two group sizes. The indication from these comparisons, therefore, is that 3-man and 4-man groups did not differ in the elicitation of the measured behaviors.

It is possible that an interaction effect was present between group size and time or stress. And any such interaction is obscured by the above analysis. However, if there are strong differences between 3-man and 4-man groups, some main effects (shown not to have occurred) would have been expected.

The product ratings in the present study (combined from 3-man and 4-man groups) matched Hackman's (3-man only) product scores very well. Had the 4-man groups differed systematically on the products, such a good match would have been extremely unlikely.

So, the data suggest that on the behaviors and performances measured in the present study, no significant differences stand between 3-man and 4-man groups. The combination of data from the two sizes of groups therefore seems legitimate.

TABLE 41

COMPARISONS OF 3-MAN AND 4-MAN GROUPS

Method	Trait	3-man Groups		4-man Groups		t	p
		Mean	Var.	Mean	Var.		
Self Ratings	Influence	4.99	0.47	5.01	0.56	0.19	ns
	Affect	5.02	0.91	5.36	0.68	1.24	ns
	Activity	4.43	0.89	4.64	0.59	0.83	ns
Peer Ratings	Influence	5.01	0.46	5.05	0.22	0.26	ns
	Affect	5.12	0.67	5.31	0.13	1.12	ns
	Activity	4.35	0.78	4.61	0.01	1.52	ns
Obs. Ratings	Influence	4.46	0.88	4.76	0.75	1.27	ns
	Affect	4.54	0.25	4.72	0.21	1.44	ns
	Activity	4.07	1.47	4.56	0.48	1.86	ns
Test (ICL)	Influence	1.16	17.27	1.48	22.13	0.24	ns
	Affect	-1.99	68.57	-1.64	53.80	0.17	ns
	Activity	30.93	110.36	27.75	126.58	0.41	ns
Event Records	Influence	0.00	0.43	0.00	0.50	0.00[a]	ns
	Affect	0.11	0.01	0.16	0.01	2.05	$\leq .05$
	Activity	1.25	0.12	1.27	0.09	0.24	ns

[a] This mean value is necessarily 0.00 for each group

VITA

John Charles Stapert

BORN: September 25, 1942, Kalamazoo, Michigan

MARITAL STATUS: Married (1 child)

EDUCATION

1948-1956	North Christian Elementary School, Kalamazoo, Michigan
1956-1960	Kalamazoo Christian High School, Kalamazoo, Michigan
1960, '61, '62 (summers only)	Western Michigan University, Kalamazoo, Michigan
1960-1963	Hope College, Holland, Michigan
1963-1966	Fuller Theological Seminary, Pasadena, California
1966-1969	University of Illinois, Urbana, Illinois

DEGREES

A.B. - 1963	Hope College, Holland, Michigan. Major: Psychology
B.D. - 1966	Fuller Theological Seminary, Pasadena, California
M.A. - 1968	University of Illinois, Urbana, Illinois. Major: Psychology
Ph.D. - 1969	University of Illinois, Urbana, Illinois. Major: Personality (Psychology) Minor: (split) Social Psychology Sociology

PUBLICATIONS

Attraction and Disagreement-Produced Arousal. With Gerald Clore, Jr. Journal of Personality and Social Psychology (in press).

EXPERIENCE

1962-1963	Student Discussion Assistant Department of Psychology, Hope College
1964-1966	Teaching Assistant in Hermeneutics Fuller Theological Seminary
1966-1968	Teaching Assistant Department of Psychology, University of Illinois
1967 (summer)	Research Assistant, University of Illinois (Dr. Frank Costin)
1968-1969	Research Assistant, University of Illinois (Dr. Joseph E. McGrath)

PROFESSIONAL AND HONORARY SOCIETIES

American Psychological Association (Professional)
Christian Association for Psychological Studies (Professional)
Eta Sigma Phi (Classics) (Honorary)
Pi Kappa Delta (Forensics) (Honorary)